Received: 5 February 2021 | Accepted: 23 August 2022

DOI: 10.1111/rssc.12601

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS
Journal of the Royal Statistical Society
Series C
Applied
Statistics

ORIGINAL ARTICLE

Optimal approximate choice designs for a two-step coffee choice, taste and choice again experiment

Nedka Dechkova Nikiforova[1] | **Rossella Berni**[1] |
Jesús Fernando López-Fidalgo[2]

[1]Department of Statistics, Computer Science, Applications "G.Parenti", University of Florence, Florence, Italy

[2]Institute of Data Science and Artificial Intelligence, University of Navarra, Pamplona, Spain

Correspondence
Nedka Dechkova Nikiforova, Department of Statistics, Computer Science, Applications "G.Parenti", University of Florence, Viale Morgagni 59, 50134 Florence, Italy.
Email: n.nikiforova@unifi.it

Funding information
Spanish Research Agency and fondos FEDER, Grant/Award Number: MTM2016-80539-C2-R1

Abstract

This work deals with consumers' preferences about coffee. Firstly, a choice experiment is performed on a sample of potential consumers. Following this, a sensory test involving the tasting of two varieties of coffee is carried out with the respondents, after which the same choice experiment is supplied to them again. An innovative approach for building heterogeneous choice designs is specifically developed for the case-study, based on approximate design theory and compound design criterion. Panel Mixed Logit models are used, thereby allowing for the inclusion of correlation among consumers' responses; choice-sets are supplied to a proportion of respondents according to optimal weights. The estimation results of the Panel Mixed Logit model are satisfactory, confirming the validity of the proposed approach.

KEYWORDS

approximate design, choice experiment, compound design criterion, correlated preferences, heterogeneous choice design, random utility models

---

1 INTRODUCTION

Starting from the seminal paper of Louviere and Woodworth (1983), choice experiments have been widely used to study and analyse the consumers' preferences for a new product or service, with applications in various fields like marketing, transportation, environmental, health and political sciences. In a choice experiment, respondents receive different choice sets for the evaluation, and are asked to choose their preferred option from each of the choice sets shown to them. Each choice-set is composed of a set of alternatives, and each alternative is a combination of the different levels of the attributes related to the characteristics of the product/service. All the alternatives submitted and included in the choice-sets form the experimental design. Respondents are assumed to be utility-maximisers, that is, within each choice-set they choose the alternative that maximises their utility, and their choices can be analysed through Random Utility Models (RUM) (McFadden, 1974). A fundamental issue when dealing with the method of choice experiments is the planning of the underlying experimental design. To this end the most widely accepted method in the literature is to build choice designs based on optimal design theory (Grasshoff et al., 2013; Kessels et al., 2006, 2009; Liu & Tang, 2015, among others).

In this paper, we deal with an innovative approach for building optimal choice designs for a non-trivial choice-experiment about coffee. Coffee is one of the most widely consumed beverages worldwide, and an important commodity in the global economy. Several international standards on coffee trading have been developed, also addressing socio-economic and environmental issues. Consumers' preferences for coffee mainly relate to product origin, social aspects and organic production, as well as to taste, aroma and health properties. They play a significant role in determining coffee consumption and promoting the sustainability of vital ecosystems, and the welfare of agricultural producers.

The case-study on coffee concerns a choice experiment, integrated with data coming from a chemical analysis of the coffee's caffeine content, and a consumer sensory test. First, two types of coffee with different organoleptic characteristics are chosen: a soft, velvety, aromatic blend (100% Arabica) and an intense, aromatic blend (Arabica and Robusta varieties). Following, the chemical analysis of the two types of coffee is performed through a High-Performance Liquid Chromatography (HPLC) evaluation method. The caffeine content obtained, opportunely standardised, is used to evaluate the consumers' preferences towards more or less strong coffee. Once the HPLC results are obtained, a consumer sensory test is also planned. For this purpose, two scoring cards (one for each type of coffee) with 10 organoleptic descriptors, are developed. Said descriptors are as follows: colour related to sight, intensity and quality of the aroma related to olfaction; four different descriptors related to taste, namely, bitter, acidic, sweet and to aroma; tactile sensation related to body, aftertaste and the general balance of the coffee. During the guided tasting, the consumers are asked to give a score for each organoleptic descriptor on a scale of seven points (Masi et al., 2013). A variable called 'Taste Score' is built, enabling evaluation of the consumers' preferences towards one of the two coffee blends. Lastly, the choice experiment is planned with six attributes at two levels: type of coffee, packaging, taste, geographical origin and certifications related to the product sustainability, and one attribute at three levels related to price. One of the main aims of the case-study is to integrate the choice experiment with the extra preference information obtained through the chemical analysis and consumers' sensory test, for including them in the corresponding design matrix.

The plan for conducting the choice experiment is as follows. First, the respondents are asked to fill in a background questionnaire containing baseline questions (e.g., age, marital status, coffee consumption and purchasing) after which they receive the choice-sets (Choice 1 before tasting). Once the first-choice experiment session is completed, the consumer sensory test is conducted by an expert who also provides information about each type of coffee and explains how to fill in the scorecards. Following, the respondents taste both types of coffee and mark their scores on the sensory assessment cards. We then supply the same choice-sets again—Choice 2 after tasting, to find out whether any differences occurred in the consumers' preferences between Choice 1 and Choice 2 sessions due to the information step and the consumer sensory test (Lombardi et al., 2017).

The procedure for conducting the choice experiment raises an issue concerning the number of choice-sets to be supplied. More precisely, a homogeneous choice design, according to which each respondent gets all the choice-sets included in the design, could produce huge respondent fatigue for the case-study, since every respondent would have to evaluate all the choice-sets contained in the design twice. To this end, we focus on heterogeneous choice designs whereby respondents receive different sub-designs, extracted from the homogeneous one. For instance, consider a choice design composed of sixteen choice-sets. In the case-study, a homogeneous choice design requires that each respondent evaluate all 16 choice-sets twice. Conversely, a heterogeneous choice design allows different respondents to evaluate a smaller number of choice-sets, for example eight, thus avoiding any overburdening of respondents. Unlike existing research in the choice design literature, we aim to build optimal heterogeneous choice designs based on an approximate design theory, instead of an exact design framework. The purpose is to obtain collections of choice-sets (i.e., sub-designs composed of a smaller number of choice-sets) to be supplied to the respondents according to associated optimal design weights, considering that the choices provided by the same respondent are correlated.

The remainder of the manuscript is organised as follows. Section 2 contains a literature review, a detailed description of the innovative contribution, also related to the case-study. In Section 3 we introduce the basic elements on optimal design theory for choice experiments and RUMs. Section 4 illustrates the innovative proposal, specifically developed for the case-study. In Section 5 we report the results for the simulation study. In Section 6 we describe in detail the case-study, also including an overall comparative assessment, and a discussion about the results. The final remarks end the manuscript.

2 | LITERATURE REVIEW AND INNOVATIVE AIMS

Various studies have already addressed the issue on how to build optimal choice designs under the most simple RUM model, that is, the Multinomial Logit (MNL), by also considering the Bayesian design framework (Kessels et al., 2006, 2009). For the MNL model, the limiting property of the Independence of Irrelevant Alternatives (IIA) is assumed (Luce, 1959; McFadden, 1974). Furthermore, the MNL model does not take account of differences in the consumers' behaviour, that is, each respondent, with different baseline characteristics is treated in the same way based only on their judgement. In the literature, the Mixed Logit model has been introduced to improve these issues (McFadden & Train, 2000). This RUM model allows for relaxing the limiting assumption of the MNL model by considering the attributes' coefficients as random variables, and not as fixed ones. In particular, depending on how the Mixed Logit model is specified, it makes it possible to account for: (i) the preference heterogeneity across consumers: the so-called Cross-Sectional Mixed Logit (C-MIXL) model (McFadden & Train, 2000), (ii) the correlation across choice observations of the same respondent: the so-called Panel Mixed Logit

(P-MIXL) model, by also allowing to evaluate the respondents' preference heterogeneity (Revelt & Train, 1998).

The construction of optimal choice designs for the C-MIXL model has been largely addressed in the literature, by considering both homogeneous designs (i.e., all the choice-sets of the design are supplied to each respondent) and heterogeneous designs (i.e., different respondents evaluate different sub-designs with a smaller number of choice-sets extracted from the homogeneous one). The first contribution about the construction of heterogeneous choice designs, and based on the exact design framework, is in Sándor and Wedel (2005). By considering Bayesian and local design framework, they built optimal heterogeneous choice designs for the MNL and C-MIXL models; the authors demonstrated a substantial improvement in the efficiency of the coefficients estimated through a heterogeneous choice design and modelling, with respect to the homogeneous one. Still considering the exact design framework, Yu et al. (2009) built Bayesian optimal homogeneous choice designs for the C-MIXL model. The main challenge in building heterogeneous choice designs through the exact design framework is that in most real situations it is computationally forbidding to do an exhaustive search in the entire design space for finding an optimal heterogeneous choice design. To this end, Liu and Tang (2015) proposed a new approach for the construction of heterogeneous choice designs for the C-MIXL model based on the approximate design theory; this approach allows for avoiding the high computational time in the computer search for heterogeneous designs, given the well-known mathematical tools for checking and guaranteeing the design optimality. Instead, for the P-MIXL model, the building of optimal choice designs has received less attention in the research literature. More precisely, by employing the exact design framework, Bliemer and Rose (2010) built optimal choice designs, while Yu et al. (2011) proposed an individually adapted sequential Bayesian approach for this purpose. Zhang et al. (2017) have recently derived a more detailed expression for the P-MIXL Information matrix and addressed the issue on how it is efficiently approximated.

More recently, Singh et al. (2019) built optimal choice designs under the MNL model with blocks of equal size, where each block was supplied to a respondent; the authors developed their approach under the utility-neutral assumption (i.e., by assuming that the unknown parameter values in the MNL Fisher Information Matrix (FIM) are all equal to 0). Singh et al. (2019) introduced an additional parameter for the block effect in the MNL model to account for a particular type of respondents' heterogeneity. Großmann (2020) extended the approach of Singh et al. (2019) by considering a MNL model with main effects only, and all attributes at two levels. The author also investigated the respondents' heterogeneity as defined by Singh et al. (2019), illustrating how it is substantially different with respect to its usual meaning in the choice experiment literature (Sándor & Wedel, 2002), since it relates to how the respondents' choices are affected by the order in which the alternatives are presented. Our approach differs from those of Singh et al. (2019) and Großmann (2020), because it relates more to a computational problem for searching optimal approximate heterogeneous choice designs. Although we are also dealing with random effects, we consider the P-MIXL model as well, since different choice-sets are evaluated by the same respondent, and therefore correlated observations are collected. Our heterogeneous choice designs do not produce different blocks, but rather, different design points with different weights (all of which are optimal in a whole), where each design point corresponds to a collection of choice-sets.

The case-study also aims to integrate the choice-experiment with the extra preference information; that is, the design matrix of the optimal heterogeneous choice design should also consider: (i) the HPLC measurement results for the coffee caffeine contents and (ii) the sensory assessment scores. To this end, we develop our proposal under a compound design criterion (Atkinson et al., 2007; Atkinson & Bogacka, 1997; Wynn, 1970) for addressing the two following main issues: i) an efficient estimation of the attributes of the choice experiment, and ii) detection of the effects related to the HPLC results (in Choice 1 modelling) and the scores obtained through the consumer sensory test (in Choice 2 modelling). As in Henderson and Liu (2016) who assume a selective choice process consisting of active and inactive attributes, a compound design criterion is used to incorporate prior information for the joint purpose of the efficient estimation of the coefficients for the active attributes, and the detection of the effects for the inactive attributes.

Motivated by the case-study, we use the approximate design framework that allows us to: i) state that our final designs are optimal since they satisfy the General Equivalence Theorem-GET (Kiefer & Wolfowitz, 1960); ii) avoid the high computation time in the search for the exact design framework. In building the optimal heterogeneous choice designs under the P-MIXL model, we consider collections of choice-sets as design points. With the aim of reducing the resulting prohibitively huge design space encountered in the case-study, we obtain a reduced design space for searching optimal collections of choice-sets under the P-MIXL model, by exploiting the optimal choice-sets obtained under the C-MIXL model. The valid use of the C-MIXL model is shown through a simulation study in which the search for the collections of choice-sets could be performed over the entire design space for the P-MIXL model. According to the simulation results, the optimal collections of choice-sets obtained by searching over the entire design space of the P-MIXL model (i.e., by considering a collection of choice-sets as a design point), contain exclusively the optimal choice-sets found under the C-MIXL model (i.e., by considering a choice-set as a design point). We would like to stress that the simulations should be distinguished from the real application, since they are only performed for supporting the validity of the proposal, applied subsequently to the case-study. Given the dependence of the C-MIXL and P-MIXL FIMs on the unknown parameter values, we consider local optimal designs, to this end, assuming nominal values for the unknown parameter values. We also perform a sensitivity analysis versus a possible wrong choice of the nominal values for the parameters, in order to provide a reference for possible loss of design efficiency. The results confirm the robustness of the optimal designs obtained under both the C-MIXL and the P-MIXL models.

3 | OUTLINE OF THE THEORY ON OPTIMAL DESIGNS FOR CHOICE EXPERIMENTS AND RUMS

In this section, we briefly illustrate the basics of approximate design theory for choice experiments and the random utility theory framework. Further details are reported in Appendix S1.

3.1 | Approximate design framework in the choice experiment context and a compound design criterion

An approximate design is represented by a probability measure ξ over a compact design space χ. In the context of choice experiments, an approximate design ξ can be expressed as (Kiefer & Wolfowitz, 1960; Liu & Tang, 2015; Tian & Yang, 2017):

$$\xi = \begin{Bmatrix} C_1 & C_2 & \dots & C_q & \dots & C_Q \\ w_1 & w_2 & \dots & w_q & \dots & w_Q \end{Bmatrix}, \tag{1}$$

where C_q is a choice-set that belongs to the space of all possible choice-sets $\chi = (C_1, \dots, C_q, \dots, C_Q)$, and w_q, $q = 1, \dots, Q$, is its corresponding weight with $0 \leq w_q \leq 1$ and $\sum_{q=1}^{Q} w_q = 1$; as stated previously, a choice-set is composed of a set of alternatives, where each alternative is a combination of different attribute levels. In an approximate design framework, the GET provides the necessary and sufficient condition for checking optimality (Kiefer & Wolfowitz, 1960); for further details on the GET we refer to Appendix S1.

In illustrating the compound design criterion used to build the optimal heterogeneous choice designs, for the moment we assume that the vector $\boldsymbol{\beta}$ of unknown coefficients to be estimated through a RUM is of dimension $K \times 1$. In order to address one of the main objectives of the case-study, we apply the following compound D-optimality criterion, $\Phi_C(\xi)$, (Atkinson et al., 2007; Atkinson & Bogacka, 1997; Wynn, 1970):

$$\Phi_C(\xi) = \frac{\alpha}{K_1}(\log|\boldsymbol{I}_{11}(\xi)|) + \frac{(1-\alpha)}{K_2}(\log|\boldsymbol{I}(\xi)| - \log|\boldsymbol{I}_{11}(\xi)|) \qquad \alpha \in [0;1], \tag{2}$$

where $K_1 + K_2 = K$. Through the criterion expressed in formula (2), we balance two objectives, namely: (i) efficient estimation of the effects of the attributes of the choice experiment (D−optimality) and (ii) detection of the effect of the HPLC results and the scores of the sensory assessment (D_s−optimality). More precisely, in formula (2), $\boldsymbol{I}_{11}(\xi)$ is the FIM containing the K_1 coefficients related to the attributes of the choice experiment, while $\boldsymbol{I}(\xi)$ is the FIM that contains both the K_1 coefficients of the choice experiment and the K_2 coefficients related to: (i) the HPLC results in Choice 1 session and (ii) the sensory assessment scores in Choice 2 session.

The coefficient α $(0 \leq \alpha \leq 1)$ reflects the relative interest in both objectives of the criterion in formula (2). When $\alpha = 1$ we obtain a D-optimal choice design for the K_1 model coefficients (i.e., for the choice experiment attributes), while for $\alpha = 0$ we obtain a D_s-optimal design for the K_2 model coefficients (i.e., for the HPLC results and the sensory assessment scores). In order to determine the best value of α, the efficiencies for a series of values of α can be calculated, and a plot of them against α offers a practical tool for choosing a design with an optimal balance for both aspects (Atkinson & Bogacka, 1997). For a design that maximises the compound criterion in formula (2), the maximum value of the directional derivative function is equal to one, providing the corresponding GET conditions (further details are reported in Appendix S1).

3.2 | Random utility framework for choice experiments and the mixed logit model

Let us assume that a respondent i $(i = 1, \dots, n)$ receives T_i choice-sets belonging to χ; for the sake of simplicity, in the rest of the theory we leave out the suffix i for T_i without any loss of generality. Let C_t be the generic tth choice-set evaluated by the respondent i, $C_t \in \chi$. According to the random utility theory framework, for each choice-set C_t, the respondent i assigns utility to each alternative j $(j = 1, \dots, J)$ expressed as (Luce, 1959; McFadden, 1974; Thurstone, 1927):

$$U_{ijt} = \boldsymbol{x}'_{ijt}\boldsymbol{\beta} + \epsilon_{ijt}. \tag{3}$$

The utility function U_{ijt}, formula (3), is composed of: (i) a systematic (observable) part, that is a deterministic function of the vector of the attribute levels $\boldsymbol{x}'_{ijt}$, while $\boldsymbol{\beta}$ is the vector of unknown coefficients to be estimated and (ii) a stochastic (unobservable) part $\boldsymbol{\epsilon}_{ijt}$ that represents the random component, assumed to be i.i.d. Gumbel or type I extreme value distributed.

The Mixed Logit model belongs to the random utility class of models (McFadden & Train, 2000); in this RUM model, the coefficients in the vector $\boldsymbol{\beta}$ are assumed as random variables, and not as fixed ones. In this paper we assume $\boldsymbol{\beta} \sim \text{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where MVN stands for the multivariate normal distribution; $\boldsymbol{\mu}$ is a vector of the means and $\boldsymbol{\Sigma}$ is a symmetric and positive definite diagonal matrix with $\sigma_1^2, \dots, \sigma_K^2$ on its main diagonal (Sándor & Wedel, 2005). Then $\boldsymbol{\beta} = \boldsymbol{\mu} + \boldsymbol{Z\sigma}$, where $\boldsymbol{Z}$ is a $K \times K$ diagonal matrix with the K elements $\boldsymbol{z} = (z_1, \dots, z_K)$ on its main diagonal, assumed to be i.i.d. standard Normal distributed. Therefore, in a Mixed Logit model, the probability that a respondent i chooses the alternative j from choice-set C_t is defined as in the following:

$$\pi_{ijt} = \int_{\mathbb{R}^K} \frac{\exp\left(\boldsymbol{x}'_{ijt}\left(\boldsymbol{\mu} + \boldsymbol{Z\sigma}\right)\right)}{\sum\limits_{j \in C_t} \exp\left(\boldsymbol{x}'_{ijt}\left(\boldsymbol{\mu} + \boldsymbol{Z\sigma}\right)\right)} \phi(z_1) \dots \phi(z_K) d\boldsymbol{z}, \tag{4}$$

where ϕ is the standard normal density function.

As already stated in Section 2, in order to develop our proposal we consider both the C-MIXL and the P-MIXL models. In a C-MIXL model, the probabilities of a respondent on a series of choice-sets are assumed to be independent. Conversely, the P-MIXL model assumes that the probabilities of a single respondent are correlated when he/she evaluates a series of choice-sets. These two assumptions lead to two different specifications of the corresponding log-likelihood functions, and therefore, to different FIMs. In our case-study, the correlation across choice observations of a single respondent is taken into account in the P-MIXL log-likelihood function, and hence, also in the P-MIXL FIM used for building the optimal designs through the compound design criterion (formula 2). The FIMs for these RUMs are reported in Appendix S1 (formulas 6 and 9). Here it is relevant to note that both FIMs have no closed form of expression since they involve multidimensional integrals that have to be approximated numerically. Furthermore, the P-MIXL FIM is more complex with respect to that of the C-MIXL one, due to the further complexity of dealing with products of logit probabilities in the log-likelihood function, compared to summations.

4 | A NEW PROPOSAL FOR THE CONSTRUCTION OF OPTIMAL HETEROGENEOUS CHOICE DESIGNS

In what follows we describe the approach specifically developed for the case-study in order to build heterogeneous choice designs, composed of collections of choice-sets, and based on an approximate design theory. We indicate such collections with $\varsigma_1, \varsigma_2, \dots, \varsigma_s, \dots, \varsigma_S$, and we assume that:

- each collection ς_s ($s = 1, \dots, S$) contains T choice-sets, that is, each respondent evaluates T choice-sets;
- no choice-set can appear twice in the same collection ς_s, but the same choice-set could appear in different collections (i.e., the T choice-sets supplied to a respondent are all different, while different respondents may evaluate the same choice-set);
- observations (i.e., the respondent's choices) within the same collection are correlated while observations between collections are independent.

Hence, we focus on an approximate design framework and generalise formula (1) in order to consider as design points the collections of choice-sets with fixed size T as follows (Atkinson & Woods, 2015):

$$\xi_S = \left\{ \begin{matrix} \varsigma_1 & \varsigma_2 & \dots & \varsigma_s & \dots & \varsigma_S \\ w_1 & w_2 & \dots & w_s & \dots & w_S \end{matrix} \right\}, \tag{5}$$

under the constraints of $0 \leq w_s \leq 1$ and $\sum_{s=1}^{S} w_s = 1$.

When considering just one choice-set as a design point, the design space χ is composed of all possible choice-sets, the total number of which is equal to Q. That is, if L is the number of all possible experimental combinations, then $Q = \binom{L}{J}$. Conversely, when the design points are the collections of choice-sets, as in formula (5), the new design space $\tilde{\chi}$ is composed of all possible collections of choice-sets of size T, the total number of which is equal to $S = \binom{Q}{T}$. In our case-study (as in most practical situations), this new design space $\tilde{\chi}$ becomes computationally forbidding. More precisely, since the choice experiment about coffee is planned with six attributes at two levels, one attribute at three levels, and two alternatives in each choice-set, then $L = 2^6 \times 3 = 192$ and $Q = \binom{192}{2} = 18336$. In the case-study we aim to build an optimal heterogeneous choice design with collections of size-eight choice-sets, by considering several pilot coffee sessions carried out previously. It follows that the design space $\tilde{\chi}$ is composed of $S = \binom{18,336}{8} = 1{,}417,526 \times 10^{39}$ possible collections of eight choice-sets. The latter clearly shows how the search is extremely burdensome even when considering approximate designs. To deal with this computational issue encountered in the case-study, we have developed the following approach in three steps, outlined in Proposal 1.

Proposal 1. Construction of optimal heterogeneous choice design for the P-MIXL model

---

1: Compute the optimal approximate choice design for the C-MIXL model, where the single design point corresponds to a choice-set, formula (1), and obtain the $2K$ optimal choice-sets with their associated design weights. Please note that we refer to a total of $2K$ optimal choice-sets since under the C-MIXL model with main effects only, we have to ensure the estimation of $2K$ coefficients.

2: Consider these $2K$ optimal choice-sets and combine them in collections of choice-sets of size T. Therefore, the reduced design space $\tilde{\chi}'$ for the search under the P-MIXL model consists of all these collections of choice-sets, the total number of which is equal to $\tilde{S} = \binom{2K}{T}$.

3: Starting from $\tilde{\chi}'$ consisting of $\tilde{S}$ collections of choice-sets (Step 2), compute the optimal heterogeneous choice design under the P-MIXL model, where each design point corresponds to a collection of choice-sets.

Once the optimal collections of choice-sets are obtained through the search over $\tilde{\chi}'$, we supply each one to a proportion of respondents according to the optimal design weights. The optimal approximate choice design is rounded off through a rounding procedure developed by Pukelsheim and Rieder (1992).

Through our proposal for the case-study rather than searching over $\tilde{\chi}$ composed of $S = \binom{18,336}{8} = 1{,}417,526 \times 10^{39}$ collections of choice-sets, we search over $\tilde{\chi}'$ composed of $\tilde{S} = 18{,}872$ collections of choice-sets, where $\tilde{S} \ll S$. To this end, we make use of both the Mixed Logit model specifications, in accordance with the underlying random utility theory related to both models. More precisely, in Step 1 we consider the C-MIXL model specification that is equivalent to the situation in which each respondent receives just one choice-set; this allows us to obtain in Step 2 the reduced design space $\tilde{\chi}'$. In Step 3, instead, we consider the P-MIXL model, that is, equivalent to the situation in which each respondent receives a subset of choice-sets.

It must be noted that through simulations we verified that the optimal collections of choice-sets obtained in Step 2 contain the same optimal choice-sets obtained in Step 1, providing strong confirmation of the proposed approach; moreover, we expect them to be near-optimal, while they may not be optimal. The simulations were carried out on a simplified choice experiment based on the case-study, in which the search for the collections in Step 2 can be performed over the entire design space $\tilde{\chi}$ (Section 5).

5 | SIMULATION RESULTS FOR APPROXIMATE HETEROGENEOUS CHOICE DESIGNS

For computing the optimal choice designs in both the simulations and the case-study, we used the modified-Optimal Weight Exchange (mOWE) algorithm (Liu & Tang, 2015), which is an extension of the Optimal Weight Exchange (OWE) algorithm proposed by Yang et al. (2013). The OWE algorithm performs very well for finding optimal approximate designs, also in terms of computation time (García-Ródenas et al., 2020; Yang et al., 2013).

For performing the simulations, we consider a simplified choice experiment with two alternatives (e.g., $J = 2$) and three attributes, one qualitative and two quantitative, based on the case-study. Therefore, we evaluate: (i) the 'Coffee Type' at two levels as the qualitative attribute and (ii) the 'Price' and the 'Taste Score', both at two levels, as the quantitative attributes. We use the compound design criterion (formula 2) for achieving an efficient estimation of the attributes 'Coffee Type' and 'Price', and detecting the effect for the attribute 'Taste Score'. Given the dependence of the FIM on the unknown coefficient values, for all the optimal designs computed here, we use the following nominal values: $\boldsymbol{\mu} = (-1.5, 1.5, -1.5)$ and $\boldsymbol{\sigma} = \left(\sqrt{0.75}, \sqrt{0.75}, \sqrt{0.75}\right)$ chosen in terms of medium response accuracy and medium respondent heterogeneity (Arora & Huber, 2001; Toubia et al., 2004; Zhang et al., 2017). All the optimal choice designs presented here are obtained through the mOWE algorithm (carried out by the Interactive Matrix Language-IML; SAS, Windows Platform vs.9.4. on a HP computer with 1.90 GHz processor speed and 8 GB RAM memory). Moreover, the design optimality is checked and verified through the GET.

5.1 | Validity by simulations: one choice-set

Following the Proposal 1, in Step 1 we first build an optimal choice design, ξ_1, under the C-MIXL model. The total number of experimental combinations, L, is equal to $L = 2^3 = 8$, and the size of all possible choice-sets Q is equal to $Q = \binom{L}{J} = \binom{8}{2} = 28$. We have to ensure the estimation of six coefficients (i.e., $2K = 6$): three coefficients related to the vector $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)$, and three coefficients related to the vector $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$. The optimal design ($\xi_1$) we obtained is supported on six design points and reported in Table 1.

In order to determine the value of α for the compound design criterion, Figure 1 shows the plot of the efficiencies against different values of α. According to this plot, we choose a value of $\alpha = 0.75$ for which we obtain 56% efficiency with respect to a D-optimal design for the 'Coffee Type' and 'Price' attributes, and 56% efficiency with respect to a D_s-optimal design for the attribute 'Taste Score'.

TABLE 1 Optimal design ξ_1

Choice-set	Coffee type	Price	Taste score	Design weights
C_4	1	−1	1	0.13
	1	−1	−1	
C_7	1	−1	−1	0.17
	−1	1	1	
C_{10}	1	1	1	0.12
	−1	1	1	
C_{17}	−1	1	1	0.21
	−1	−1	1	
C_{20}	1	1	−1	0.18
	−1	−1	1	
C_{21}	1	1	1	0.19
	−1	−1	1	

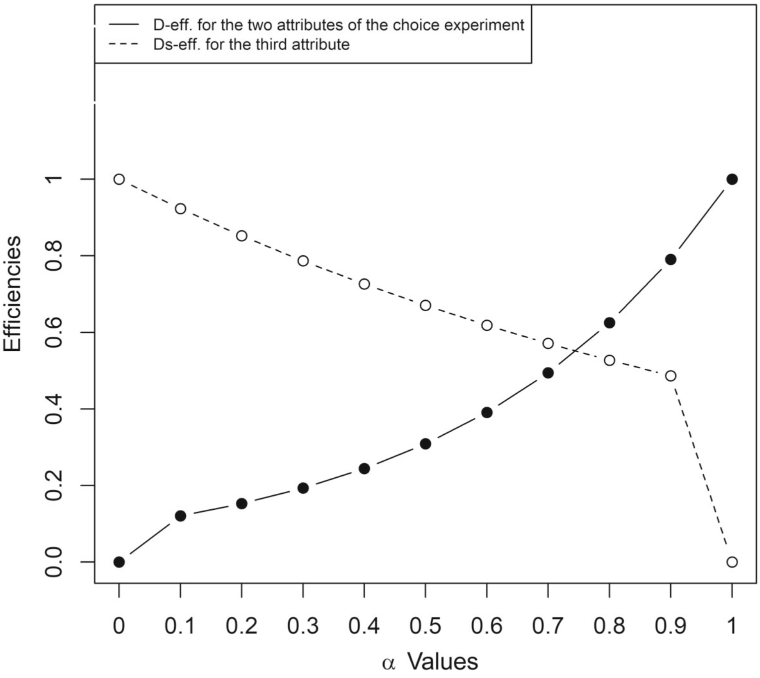

FIGURE 1 Efficiencies for different values of α for design ξ_1

5.2 | Validity by simulations: collections of two, three, four and five choice-sets

Once we have obtained the optimal choice design ξ_1 at Step 1 (Proposal 1), we proceed to compute the optimal heterogeneous choice designs composed of collections of two, three, four and five choice-sets (Steps 2 and 3, Proposal 1). When considering the collections of choice-sets as design points, the size of the design space increases exponentially when the number of choice-sets in each collection increases. In fact, the design space $\tilde{\chi}$ rises from $\binom{Q}{S} = \binom{28}{2} = 378$ for collections with two choice-sets, to $\binom{Q}{T} = \binom{28}{5} = 98{,}280$ for collections consisting of five choice-sets.

Table 2 contains the optimal heterogeneous choice designs obtained, all of which are checked via GET for collections of two to five choice-sets (labelled ξ_2, ξ_3, ξ_4 and ξ_5, respectively). It must be noted that the numbering of the choice-sets composing the optimal collections is obtained computationally through the function 'allcomb' in SAS, Interactive Matrix Language-IML (SAS, Windows Platform vs.9.4.). The optimal designs found under the P-MIXL model with a collection of choice-sets as a design point (ξ_2, ξ_3, ξ_4 and ξ_5, Table 2), contain exclusively the optimal choice-sets selected in the design ξ_1 found under the C-MIXL model (Table 1). These results play a fundamental role for validating our proposal; in fact, they show that starting from the optimal choice-sets for the C-MIXL model, we can obtain an optimal choice design for the P-MIXL model. The computation time for obtaining the optimal designs is very fast (Table 3), given the use of the mOWE algorithm (Liu & Tang, 2015; Yang et al., 2013).

In Figure 2 we report the efficiencies of the optimal choice designs (ξ_2, ξ_3, ξ_4 and ξ_5) against different values of α. As can be observed from the plots in Figure 2, there is an increasing efficiency when the number of choice-sets per collection increases. In fact, the efficiency increases from 66% for collections with two choice-sets to 76% for collections composed of five choice-sets. However, there is no mathematical solution which could unequivocally indicate the ideal number of choice-sets per collection. Instead, this feature seems to be strictly related to the specific problem under study (Chung et al., 2010).

TABLE 2 Optimal heterogeneous choice designs for collections of two, three, four and five choice-sets

Design	Design points	Design weights
ξ_2		
	(C_7, C_{10})	0.31
	(C_{20}, C_{21})	0.32
	(C_{10}, C_{21})	0.01
	(C_4, C_{17})	0.33
	(C_7, C_{21})	0.03
ξ_3		
	(C_4, C_7, C_{10})	0.17
	(C_7, C_{10}, C_{17})	0.33
	(C_4, C_{20}, C_{21})	0.32
	(C_{17}, C_{20}, C_{21})	0.18
ξ_4		
	$(C_4, C_7, C_{10}, C_{20})$	0.35
	$(C_4, C_{10}, C_{17}, C_{21})$	0.36
	$(C_7, C_{17}, C_{20}, C_{21})$	0.29
ξ_5		
	$(C_4, C_7, C_{10}, C_{17}, C_{21})$	0.37
	$(C_4, C_{10}, C_{17}, C_{20}, C_{21})$	0.34
	$(C_4, C_7, C_{10}, C_{17}, C_{20})$	0.29

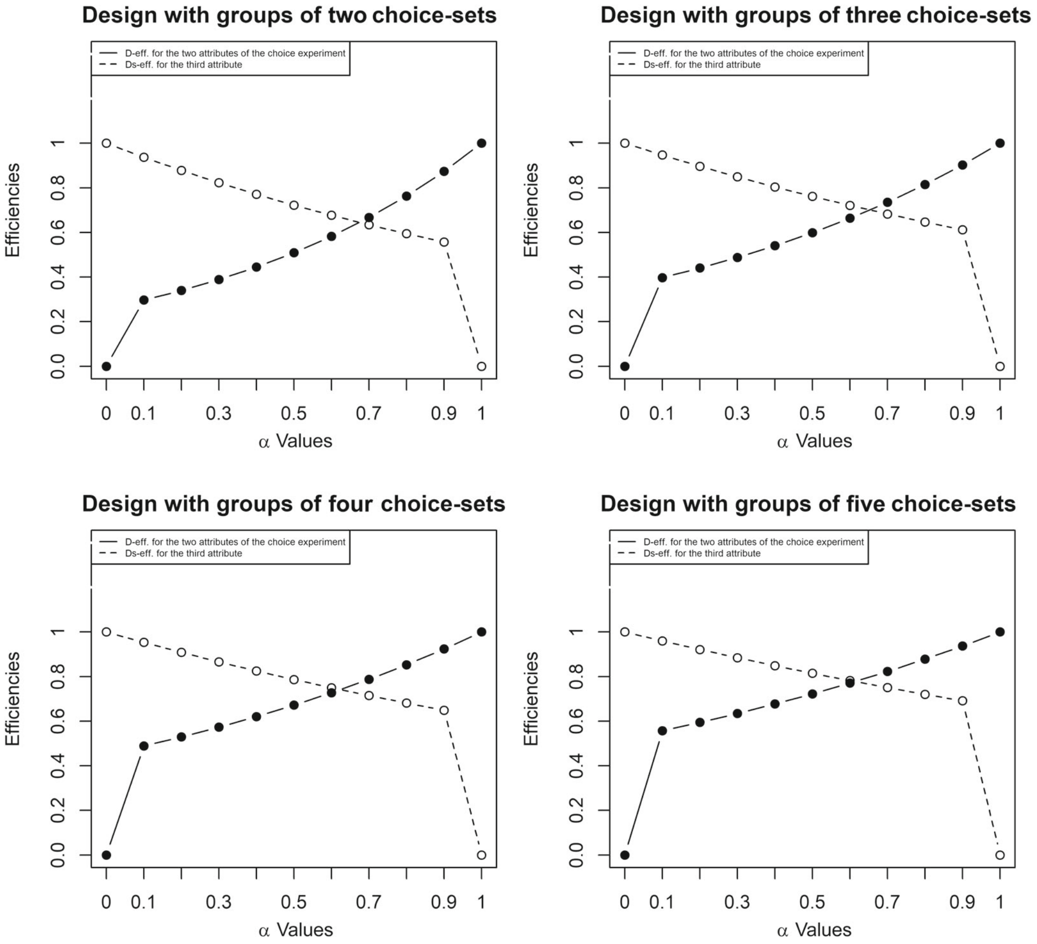

FIGURE 2 Efficiencies for different values of α for designs ξ_2, ξ_3, ξ_4 and ξ_5

TABLE 3 Computation time for the optimal designs

Optimal design	Computation time (mm:ss)
ξ_1	01:03
ξ_2	01:06
ξ_3	01:10
ξ_4	02:32
ξ_5	08:54

6 | A REAL CASE-STUDY

6.1 | Attributes and levels for the choice experiment, chemical analysis and consumers sensory test

We identified six attributes of the choice experiment, chosen after a thorough investigation of the possible types of coffee available from mass market retailers, with specific focus on types of coffee, sensory properties and price. The first is the type of coffee, labelled by ‘Coffee Type’, at two levels (a blend of Arabica and Robusta and a blend of 100% Arabica); the second is the packaging (labelled by ‘Packaging’) at two levels: a soft bag and a jar, both in a modified atmosphere. We established two attributes related to the taste: a Soft and Velvety taste (‘Soft Velvety Taste’) and an Intense and Aromatic taste (‘Intense Aromatic Taste’), both at two levels, for example, fairly present and highly present. Moreover, in order to evaluate the consumers’ preferences regarding the sustainability and geographical origin, we included the attribute ‘Label Indication’ at two levels: ‘Geographical Origin’, that is, the indication of the geographical origin of the coffee, and ‘Certification of Sustainability’, that is, the presence of any type of certification of sustainability (economic, social and/or environmental). The last attribute is the price (‘Price’) for a quantity of 250 g of coffee at three levels: €4.50, €6.00 and €7.50. The attributes are reported in Table 4, with correspondent coded levels.

In planning the choice experiment, we chose to build choice-sets with two alternatives, and to supply a collection of eight choice-sets to each respondent by considering: (i) the attributes with their corresponding levels, selected after a thorough investigation, also with the aid of coffee experts, (ii) the non-trivial procedure for supplying the choice experiment, which implies a substantial respondent burden and (iii) the pilot coffee sessions carried out previously.

The caffeine contents obtained through the HPLC analysis registered an average value of 53.67 and 81.08 mg/cup for the 100% Arabica and the Arabica and Robusta blend, respectively. According to the HPLC results, more caffeine means a blend of Arabica and Robusta varieties, and vice versa. The variable ‘Taste Score’ is a synthesis of the scores assigned by each respondent during the guided tasting session. This variable is evaluated within the experimental design in the range [−1, 1], where the extreme value (−1) is related to Arabica and Robusta blend, while the extreme value (+1) is related to 100% Arabica blend. Thus, the more the ‘Taste Score’ goes towards −1, the more the consumer’s preferences tends towards the Arabica and Robusta blend, and vice versa.

TABLE 4 Attributes and levels for the choice experiment

Attribute	Original levels (coding in brackets)
Coffee type	1: Blend of Arabica and Robusta (−1)
	2: Blend of 100% Arabica (1)
Packaging	1: Soft bag in a modified atmosphere (−1)
	2: Jar in a modified atmosphere (1)
Label Indication	1: Geographical origin (−1)
	2: Certification of sustainability (1)
Intense Aromatic Taste	1: Fairly present (−1)
	2: Highly present (1)
Soft Velvety Taste	1: Fairly present (−1)
	2: Highly present (1)
Price	1: €4.50
	2: €6.00
	3: €7.50

6.2 | Optimal heterogeneous choice designs

In this subsection we describe the building of the design matrix for our real case-study by applying our proposal (Proposal 1, Section 4). Therefore, first we obtain the optimal choice design under the C-MIXL model (Step 1, Proposal 1, Section 4). Afterwards, following Steps 2 and 3 of Proposal 1 (Section 4), we build the optimal heterogeneous choice design composed of collections of eight choice-sets. We have six attributes related to the choice experiment, and one attribute related to the caffeine results from the HPLC analysis (evaluated in Choice 1 modelling), and the sensory scores (evaluated in Choice 2 modelling) at two coded levels; namely, the low level (-1) related to the caffeine results and the scores for the blend Arabica and Robusta, and the high level (1) related to those obtained for the blend 100% Arabica. It must be noted that the estimated P-MIXL model for Choice 1 session includes the attribute 'Caffeine' at two levels related to the quantity of caffeine in the model. Instead, the estimated P-MIXL model for Choice 2 session includes the attribute 'Taste Score' related to the scores collected through the guided tasting.

6.2.1 | Optimal choice design when each respondent receives one choice-set

By following our proposal (Section 4), firstly (Step 1, Proposal 1) we build the optimal heterogeneous choice design under the C-MIXL model. The total number of experimental combinations is 192, and by considering two alternatives, the total number of possible choice-sets is 18,336. We have to estimate sixteen coefficients: eight related to the vector $\boldsymbol{\mu} = (\mu_1, \dots, \mu_8)$ and eight related to the vector $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_8)$. As before, we used the following nominal values (Arora & Huber, 2001; Toubia et al., 2004; Zhang et al., 2017): $\boldsymbol{\mu} = (-1.5, 1.5, -1.5, 1.5, -1.5, 1.5, 0, 1.5)$ and $\boldsymbol{\sigma} = (\sqrt{0.75}, \sqrt{0.75}, \sqrt{0.75}, \sqrt{0.75}, \sqrt{0.75}, \sqrt{0.75}, \sqrt{0.75}, \sqrt{0.75})$. The optimal choice design we obtained under the C-MIXL model is supported on 16 points as reported in Table 5, and the design optimality is checked via GET.

By plotting the efficiencies for different values of α in Figure 3 (left panel), a reasonable value of α is equal to 0.72 for which we obtain 57% efficiency with respect to a D-optimal design for the choice experiment attributes, and 57% efficiency with respect to a D_s-optimal design related to the caffeine and the sensory assessment scores.

6.2.2 | Optimal heterogeneous choice design when each respondent receives a collection of eight choice-sets

In line with our proposal, we compute the optimal heterogeneous choice design composed of collections of eight choice-sets (Steps 2 and 3, Proposal 1, Section 4). The optimal heterogeneous choice-design obtained is supported on six design points reported in Table 6. From plotting the efficiencies for different values of α in Figure 3 (right panel), a reasonable value of α is equal to 0.8 for which we obtain 75% efficiency with respect to a D-optimal design for the choice experiment attributes, and 75% efficiency with respect to a D_s-optimal design related the attribute of 'Caffeine' (Choice 1) and 'Taste Score' (Choice 2). In Table 7 we report the computation time for obtaining the optimal designs: the results show a very good overall performance for both designs ξ^* and $\xi^*_{\tilde{S}}$.

TABLE 5 Optimal design ξ^*

Choice-set	Intense aromatic taste	Packaging	Soft velvety taste	Label indication	Coffee type	Vector 1 for price	Vector 2 for price	Taste score	Design weights
C_{1423}	1	1	1	1	−1	0	1	1	0.06
	1	−1	1	1	1	−1	−1	−1	
C_{3337}	1	−1	1	−1	−1	1	0	1	0.01
	1	−1	−1	1	−1	−1	−1	1	
C_{3494}	1	−1	−1	1	1	−1	−1	1	0.04
	1	−1	−1	−1	1	1	0	1	
C_{3955}	1	−1	1	1	1	−1	−1	1	0.05
	1	−1	−1	−1	1	0	1	−1	
C_{6903}	1	1	1	1	1	0	1	1	0.06
	−1	1	1	−1	−1	1	0	−1	
C_{8328}	1	−1	1	1	−1	0	1	−1	0.07
	−1	1	−1	1	−1	0	1	−1	
C_{8857}	1	−1	1	1	−1	1	0	1	0.09
	−1	1	−1	−1	1	1	0	−1	
$C_{11,050}$	−1	1	−1	1	1	−1	−1	−1	0.09
	−1	−1	1	1	1	−1	−1	−1	
$C_{11,647}$	−1	1	−1	−1	1	0	1	1	0.13
	−1	−1	1	1	−1	0	1	−1	
$C_{12,388}$	1	1	1	−1	1	0	1	−1	0.05
	−1	−1	1	−1	1	1	0	−1	
C_{12939}	−1	1	1	1	−1	1	0	1	0.04
	−1	−1	1	−1	1	−1	−1	−1	
$C_{13,601}$	1	−1	−1	−1	−1	−1	−1	1	0.14
	−1	−1	1	−1	−1	0	1	−1	
$C_{16,646}$	1	1	1	1	−1	1	0	−1	0.03
	−1	−1	−1	−1	1	0	1	1	
$C_{16,768}$	1	−1	1	−1	−1	0	1	1	0.04
	−1	−1	−1	−1	1	0	1	−1	
$C_{16,852}$	−1	−1	−1	1	1	−1	−1	1	0.06
	−1	−1	−1	−1	1	0	1	1	
$C_{17,122}$	1	−1	−1	1	−1	−1	−1	−1	0.04
	−1	−1	−1	−1	1	−1	−1	−1	

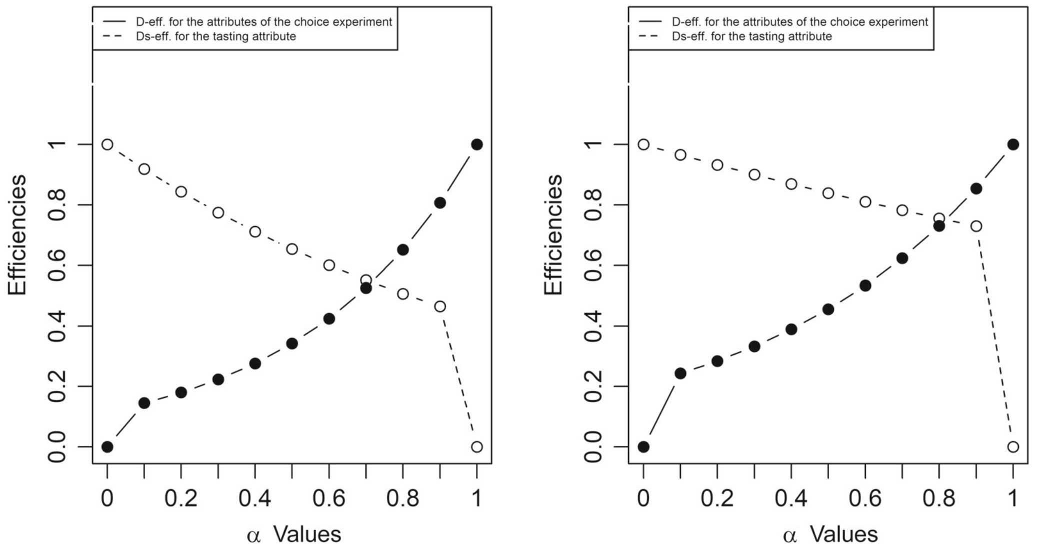

FIGURE 3 Efficiencies for different values of α for designs ξ^* (left panel) and $\xi^*_{\tilde{S}}$ (right panel)

TABLE 6 Optimal design $\xi^*_{\tilde{S}}$

Design points	Design weights
$\{C_{12388}, C_{13601}, C_{17122}, C_{8328}, C_{3494}, C_{1423}, C_{16,852}, C_{16,768}\}$	0.10
$\{C_{3955}, C_{8857}, C_{3494}, C_{6903}, C_{11647}, C_{11050}, C_{16,768}, C_{3337}\}$	0.15
$\{C_{3955}, C_{8857}, C_{13601}, C_{8328}, C_{3494}, C_{11,647}, C_{12,939}, C_{3337}\}$	0.13
$\{C_{17,122}, C_{3494}, C_{1423}, C_{16,646}, C_{11,647}, C_{11,050}, C_{12,939}, C_{16,768}\}$	0.14
$\{C_{3955}, C_{12388}, C_{13,601}, C_{1423}, C_{16,646}, C_{11,647}, C_{16,852}, C_{12,939}\}$	0.25
$\{C_{12,388}, C_{8857}, C_{17,122}, C_{8328}, C_{6903}, C_{16,646}, C_{11,050}, C_{3337}\}$	0.23

TABLE 7 Computation time for the optimal designs

Optimal design	Computation time (mm:ss)
ξ^*	04:40
$\xi^*_{\tilde{S}}$	23:43

6.3 | P-MIXL models' results for the case-study

In what follows, we describe the estimation results obtained for the P-MIXL model related to the data collected for a total of 51 respondents by considering Choice 1 and Choice 2; for each kth attribute ($K = 8$) we estimate its main effect labelled by μ_k, and its heterogeneity effect labelled by σ_k, the results are reported in Tables 8 and 9, respectively. Moreover, each $\hat{\sigma}_k$ estimate expresses the effect related to the respondents' heterogeneity, that is, for a given attribute k, the higher the value $\hat{\sigma}_k$ estimates, the greater the consumers' heterogeneity for the kth attribute. Both models are estimated by applying the *mixlogit* command of the STATA software (Hole, 2017).

TABLE 8 Panel mixed logit model results before tasting: Choice 1

	Estimate	SE	*Z* value	*p*-Value
Main effect (μ_k)				
Coffee type	0.3810	0.2176	1.75	0.0800
Packaging	−0.8500	0.2771	−3.07	0.0020
Intense aromatic taste	−0.4603	0.2138	−2.15	0.0310
Soft velvety taste	0.7570	0.2707	2.80	0.0050
Label indication	0.4486	0.1855	2.42	0.0160
Caffeine	−2.0113	0.5078	−3.96	0.0000
Price 1 (level 1)	0.0000	.	.	.
Price 2 (level 2)	−0.9297	0.4488	−2.07	0.0380
Price 3 (level 3)	−1.6424	0.4989	−3.29	0.0010
Heterogeneity effect (σ_k)				
Coffee type	0.4983	0.2151	2.32	0.0200
Packaging	1.1689	0.2745	4.26	0.0000
Intense aromatic taste	0.1495	0.2369	0.63	0.5280
Soft velvety taste	0.8737	0.3035	2.88	0.0040
Label indication	0.2680	0.1728	1.55	0.1210
Caffeine	−0.7919	0.4546	−1.74	0.0810
Price 1 (level 1)	0.0000	.	.	.
Price 2 (level 2)	1.0632	0.5934	1.79	0.0730
Price 3 (level 3)	−1.1070	0.4577	−2.42	0.0160

When observing the estimation results for Choice 1 modelling (Table 8), the positive sign of the estimated coefficient related to the 'Coffee Type' indicates a preference for the 100% Arabica blend. This result is also confirmed when considering the estimated coefficient related to the Soft and Velvety taste: its positive sign indicates a preference for its high presence that is typical of the 100% Arabica blend. Moreover, for the Intense and Aromatic taste, its negative sign indicates a preference for its fair presence; a result that is in line with the 100% Arabica blend chosen. However, the negative sign of the 'Caffeine', related to the HPLC results, indicates a preference for the Arabica and Robusta blend. Moreover, the negative sign of the 'Packaging' indicates a preference for the soft bag with a modified atmosphere with respect to the jar in a modified atmosphere. When considering the 'Label Indication', its positive sign indicates that the consumers choose the certification of product sustainability with respect to the indication of geographical origin. Lastly, the second and the third price levels are both negative with respect to the reference level (i.e., the first price level); thus, the consumers' willingness to pay decreases when price increases (as expected).

Instead, when considering Choice 2 modelling, we can observe a relevant change in the consumers' preferences with respect to Choice 1 modelling. More specifically, the negative sign related to the 'Coffee Type' indicates a preference for the Arabica and Robusta blend. Coherently with this result, the positive sign of the Intense and Aromatic taste indicates a preference for

TABLE 9 Panel mixed logit model results after tasting: Choice 2

	Estimate	SE	Z value	*p*-Value
Main effect (μ_k)				
Coffee type	−0.8493	0.2901	−2.93	0.0030
Packaging	−0.7935	0.2543	−3.12	0.0020
Intense aromatic taste	0.3803	0.1767	2.15	0.0310
Soft velvety taste	−0.4144	0.1627	−2.55	0.0110
Label indication	0.8008	0.2258	3.55	0.0000
Taste score	−2.5042	0.9690	−2.58	0.0100
Price 1 (level 1)	0.0000	.	.	.
Price 2 (level 2)	−1.3133	0.4068	−3.23	0.0010
Price 3 (level 3)	−1.7887	0.5526	−3.24	0.0010
Heterogeneity effect (σ_k)				
Coffee type	0.9126	0.2859	3.19	0.0010
Packaging	0.8528	0.2661	3.21	0.0010
Intense aromatic taste	0.5062	0.2705	1.87	0.0610
Soft velvety taste	0.3148	0.1903	1.65	0.0980
Label indication	−0.4199	0.1959	−2.14	0.0320
Taste score	3.3055	1.2044	2.74	0.0060
Price 1 (level 1)	0.0000	.	.	.
Price 2 (level 2)	0.3833	0.3699	1.04	0.3000
Price 3 (level 3)	0.1227	0.4044	0.30	0.7620

its high presence level (typical for the Arabica and Robusta blend); moreover, in line with this last result and also with the coffee type preference, the negative sign of the soft and velvety taste coefficient confirms the consumers preference towards its fair presence level. Furthermore, the negative sign related to the ‘Taste Score’, in accordance with the preference expressed for the ‘Coffee Type’, indicates a preference for the Arabica and Robusta blend. This result confirms the fact that the guided tasting, together with the information step, makes the consumers more capable of better differentiating between the two types of coffee and also of discriminating between different tastes, for example, a Soft and Velvety taste and an Intense and Aromatic taste. By considering the packaging, the negative sign of its estimated coefficient confirms the consumers preference for the soft bag in modified atmosphere. Similarly, the positive sign of the ‘Label Indication’ estimated coefficient confirms the preference for sustainability with respect to the indication of geographical origin. Lastly, the estimated coefficients related to the price attribute show the same pattern obtained in Choice 1 modelling, confirming the willingness-to-pay results obtained in Choice 1 modelling.

When considering the estimates related to the heterogeneity vector $\boldsymbol{\sigma}$, it can be observed how most of them are statistically significant in both Choice 1 and Choice 2 modelling. More precisely, in Choice 1 modelling, except for the intense and aromatic taste and the ‘Label Indication’, all the estimated coefficients related to the heterogeneity effect are statistically significant, indicating a large preference variation among individuals. Conversely, in Choice 2 modelling the coefficient related to the 'Taste Score', even though with a high heterogeneity effect, confirms the fact that after the guided tasting session the consumers' preferences are better defined.

6.3.1 | An overall comparative assessment in a dynamic choice modelling framework

In this subsection we further extend the results previously presented (Section 6.3), by integrating them with a brief analysis, open to further developments. More precisely, we aim to evaluate the consumers' preferences expressed in both sessions through only one P-MIXL model, and including a specific parameter in order to evaluate the comparison between the two sessions. In the literature, we can find some similar situations. For instance, Tully et al. (2020) compare the results between two countries, by considering pooled data, and first-order interactions involving attributes and a specific dummy variable for the respondents' nationality. In our setting, we do

TABLE 10 Panel mixed logit model results: an overall comparative assessment

	Estimate	SE	*Z* value	*p*-Value
Main effect (μ_k)				
Coffee Type	−0.8937	0.3162	−2.83	0.0050
Packaging	−0.8649	0.2606	−3.32	0.0010
Intense Aromatic Taste	0.3729	0.1747	2.13	0.0330
Soft Velvety Taste	−0.3846	0.1615	−2.38	0.0170
Label Indication	0.8937	0.2567	3.48	0.0000
Taste Score	−2.5435	0.8761	−2.90	0.0040
Price 1 (level 1)	0.0000	.	.	.
Price 2 (level 2)	−1.4320	0.4069	−3.52	0.0000
Price 3 (level 3)	−2.0682	0.6157	−3.36	0.0010
CS_1	−0.3736	0.2217	−1.68	0.0920
Heterogeneity Effect (σ_k)				
Coffee Type	1.0411	0.3195	3.26	0.0010
Packaging	0.9195	0.2846	3.23	0.0010
Intense Aromatic Taste	0.4961	0.2737	1.81	0.0700
Soft Velvety Taste	0.3349	0.1836	1.82	0.0680
Label Indication	−0.3834	0.2105	−1.82	0.0690
Taste Score	3.1624	1.1802	2.68	0.0070
Price 1 (level 1)	0.0000	.	.	.
Price 2 (level 2)	0.3651	0.3745	0.97	0.3300
Price 3 (level 3)	−0.0457	0.4086	−0.11	0.9110
CS_1	0.4707	0.2895	1.63	0.1040

not deal with two independent groups of respondents; rather, the same respondent faces the same choice experiment in two different time occasions (i.e. in Choice 1 and in Choice 2 sessions). To this end, by referring to a dynamic choice modelling framework (Heckman, 1981; Train, 2003), we include a specific variable in the P-MIXL model for Choice 2 session, that is the response variable obtained in Choice 1 session. Such variable, hereinafter labelled by 'CS_1', allows for evaluating if there is a discrepancy (i.e., a change) in the consumers' preferences in Choice 2 session with respect to Choice 1, that could be probably due to the information step and the guided tasting.

When observing the estimation results for this additional P-MIXL model (Table 10), the negative sign of the estimated coefficient related to 'CS_1' indicates that there is a discrepancy, that is, a change in the consumers' preferences between Choice 1 and Choice 2 sessions; it should be also noted that this estimated coefficient is statistically significant at the 10% level. In fact, as already described in the previous section, the consumers change their preferences for some of the attributes after experiencing the information step, and the guided tasting. It is also relevant to note that the signs of the estimated coefficients for all the attributes (Table 10) are exactly the same as those obtained for the P-MIXL model reported in Table 9, hence fully confirming the coherency of the results obtained.

6.4 | Discussion

When considering the evidences emerging from the real case-study, we note that the guided tasting session, together with the information provided on each type of coffee, plays a relevant role in unequivocally determining the respondents' preferences. As a matter of fact, when facing the Choice 1 session, the consumer expresses his/her preference only considering his/her previous knowledge about the coffee and own lifestyle; this key-point probably explains the consumers' preference for the 100% Arabica blend jointly with a preference for more caffeine in Choice 1 modelling. This initial and controversial result is however solved when facing Choice 2 session where the respondents were exposed to: (i) information given by an expert about the taste properties and organoleptic characteristics of the two types of coffee, and (ii) a guided tasting of the two types of coffee carried out by the same expert. Both the information provided by the expert and the guided tasting contribute for better defining the consumers' preferences in Choice 2 modelling with respect to Choice 1. In fact, as a further confirmation of this result, the change between Choice 1 and Choice 2 sessions, when considering the consumers' preferences, relates to the type of coffee and the two attributes related to the coffee taste only. Furthermore, the estimated 'Taste Score' coefficient in Choice 2 modelling further confirms the relevant role of the guided tasting for better defining the respondents' preferences. Lastly, the additional estimated P-MIXL model further confirms all of these results (Section 6.3.1).

It must also be noted that, for both P-MIXL models, the convergence is perfectly achieved, and we also obtained significant p-values related to the Likelihood-Ratio test (p-value = 0.0002 and 0.0129 for Choice 1 and Choice 2 modelling, respectively). The same also apply when considering the P-MIXL model presented in Section 6.3.1, for which a significant p-value related to the Likelihood-Ratio test is also obtained (p-value = 0.0082). Moreover, almost all the estimated coefficients for Choice 1 and Choice 2 are statistically significant with small standard errors. All these results further confirm the efficiency of the underlying choice design, and consequently the overall validity of our innovative approach for the case-study.

6.5 | Final remarks

In this article we proposed an innovative approach for the construction of optimal heterogeneous choice designs for correlated preferences, and under a compound design criterion, motivated by the case-study about coffee consumption. Our proposal exploits an approximate design theory that allows us to: (i) obtain heterogeneous choice designs consisting of collections of choice-sets to be supplied to the respondents according to the associated design weights and (ii) verify that the final choice designs are really optimal through the GET. The case-study show coherent results, also confirming that the proposal is suitable for applications in other real settings or environments.

Although we believe that we developed a substantial improvement in generating heterogeneous choice designs, we are fully aware that our proposal is not without limitations. More precisely, given the dependence of the FIM on the unknown coefficient values, and similarly to Sándor and Wedel (2002) and Liu and Tang (2015), we assumed nominal values for obtaining the optimal choice designs, also verified by the sensitivity analysis. However, further analyses could be performed through a broadening of the proposal to include the Bayesian design framework and an extension of the case-study, and a joint investigation of the results of the design efficiency, number of choice-set collections, and associated design weights.

ACKNOWLEDGEMENTS

The third author have been sponsored by the Spanish Research Agency and fondos FEDER MTM2016-80539-C2-R1. The authors thank Ms. Susan Mary Cadby for her revision of the English language aspects of the paper. Open Access Funding provided by Universita degli Studi di Firenze within the CRUI-CARE Agreement.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Nedka Dechkova Nikiforova https://orcid.org/0000-0002-5948-5917
Rossella Berni https://orcid.org/0000-0002-7782-1777
Jesús Fernando López-Fidalgo https://orcid.org/0000-0001-7502-8188

REFERENCES

Arora, N. & Huber, J. (2001) Improving parameter estimates and model prediction by aggregate customization in choice experiments. *Journal of Consumer Research*, 28(2), 273–283.

Atkinson, A.C. & Bogacka, B. (1997) Compound $D-$ and D_s- optimum designs for determining the order of a chemical reaction. *Technometrics*, 39(4), 347–356.

Atkinson, A.C., Donev, A.N. & Tobias, R.D. (2007) *Optimum experimental designs, with SAS*. New York: Oxford University Press.

Atkinson, A.C. & Woods, D.C. (2015) Ch. 13. Designs for generalized linear models. In: Dean, A., Morris, M., Stufken, J. & Bingham, D. (Eds.) *Handbook of design and analysis of experiments*. Boca Raton: Chapman & Hall, pp. 471–548.

Bliemer, M.C.J. & Rose, J.M. (2010) Construction of experimental designs for mixed logit models allowing for correlation across choice observations. *Transportation Research Part B: Methodological*, 44, 720–734.

Chung, C., Boyer, T. & Han, S. (2010) How many choice sets and alternatives are optimal? Consistency in choice experiments. *Agribusiness*, 27, 114–125.

García-Ródenas, R., García-García, J.C., López-Fidalgo, J., Martín-Baos, J.A. & Wong, W.K. (2020) A comparison of general-purpose optimization algorithms for finding optimal approximate experimental designs. *Computational Statistics and Data Analysis*, 144, 106844.

Graßhoff, U., Großmann, H., Holling, H. & Schwabe, R. (2013) Optimal design for discrete choice experiments. *Journal of Statistical Planning and Inference*, 143, 167–175.

Großmann, H. (2020) On the meaning of block effects in paired comparison choice experiments and a relationship with blocked 2^K main effects plans. *Journal of Statistical Planning and Inference*, 209, 76–84.

Heckman, J.J. (1981) *Heterogeneity and state dependence*. In: Rosen, S. (Ed.) *Studies in labor markets*. Chicago: University of Chicago Press, pp. 91–140.

Henderson, T. & Liu, Q. (2016) Efficient design and analysis for a selective choice process. *Journal of Marketing Research*, 54(3), 430–446.

Hole, A.R. (2017) Fitting mixed logit models by using maximum simulated likelihood. *The Stata Journal*, 7(3), 388–401.

Kessels, R., Goos, P. & Vanderbroek, M. (2006) A comparison of criteria to design efficient choice experiments. *Journal of Marketing Research*, 43, 409–419.

Kessels, R., Jones, B., Goos, P. & Vandebroek, M. (2009) An efficient algorithm for constructing Bayesian optimal choice designs. *Journal of Business & Economic Statistics*, 27(2), 279–291.

Kiefer, J. & Wolfowitz, J. (1960) The equivalence of two extremum problems. *Canadian Journal of Mathematics*, 12, 363–366.

Liu, Q. & Tang, Y. (2015) Construction of heterogeneous conjoint choice designs: a new approach. *Marketing Science*, 34(3), 346–366.

Lombardi, G.V., Berni, R. & Rocchi, B. (2017) Environmental friendly food. Choice experiment to assess consumer's attitude toward climate neutral milk: the role of communication. *Journal of Cleaner Production*, 142, 257–262.

Louviere, J.J. & Woodworth, G. (1983) Design and analysis of simulated consumer choice or allocation experiments: an approach based on aggregate data. *Journal of Marketing Research*, 20, 350–367.

Luce, R. (1959) *Individual choice behavior: a theoretical analysis*. New York: Wiley.

Masi, C., Dinnella, C., Barnabà, M., Navarini, L. & Monteleone, E. (2013) Sensory properties of under-roasted coffee beverages. *Journal of Food Science*, 78(8), 1290–1300.

McFadden, D. (1974) Conditional logit analysis of qualitative choice behavior. In: Zarembka, P. (Ed.) *Frontiers in econometrics*. New York: Academic Press, pp. 105–142.

McFadden, D. & Train, K. (2000) Mixed MNL models for discrete response. *Journal of Applied Econometrics*, 15(5), 447–470.

Pukelsheim, F. & Rieder, S. (1992) Efficient rounding of approximate designs. *Biometrika*, 79(4), 763–770.

Revelt, D. & Train, K. (1998) Mixed Logit with repeated choices: households' choices of appliance efficiency level. *Review of Economics and Statistics*, 80, 1–11.

Sándor, Z. & Wedel, M. (2002) Profile construction in experimental choice designs for mixed logit models. *Marketing Science*, 21(4), 455–475.

Sándor, Z. & Wedel, M. (2005) Heterogeneous conjoint choice designs. *Journal of Marketing Research*, 42(2), 210–218.

Singh, R., Das, A. & Chai, F. (2019) Optimal paired choice block designs. *Statistica Sinica*, 29(3), 1419–1438.

Thurstone, L.L. (1927) The law of comparative judgment. *Psychological Review*, 34, 273–286.

Tian, T. & Yang, M. (2017) Efficiency of the coordinate-exchange algorithm in constructing exact optimal discrete choice experiments. *Journal of Statistical Theory and Practice*, 11(2), 254–268.

Toubia, O., Hauser, J.R. & Simester, D.I. (2004) Polyhedral methods for adaptive choice-based conjoint analysis. *Journal of Marketing Research*, 41(1), 116–131.

Train, K. (2003) *Discrete choice methods with simulation*. New York: Cambridge University Press.

Tully, M.P., Bernsten, C., Aitken, M. & Vass, C. (2020) Public preferences regarding data linkage for research: a discrete choice experiment comparing Scotland and Sweden. *BMC Medical Informatics and Decision Making*, 20(1), 109–122.

Wynn, H.P. (1970) The sequential generation of D-optimal experimental designs. *Annals of Mathematical Statistics*, 41, 1055–1064.

Yang, M., Biedermann, S. & Tang, E. (2013) On optimal designs for nonlinear models: a general and efficient algorithm. *Journal of the American Statistical Association*, 108, 1411–1420.

Yu, J., Goos, P. & Vandebroek, M. (2009) Efficient conjoint choice designs in the presence of respondent heterogeneity. *Marketing Science*, 28(1), 122–135.

Yu, J., Goos, P. & Vandebroek, M. (2011) Individually adapted sequential Bayesian designs for conjoint choice experiments. *International Journal of Research in Marketing*, 28(4), 378–388.

Zhang, W., Mandal, A. & Stufken, J. (2017) Approximations of the information matrix for a panel mixed logit model. *Journal of Statistical Theory and Practice*, 11(2), 269–295.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Nikiforova, N.D., Berni, R. & López-Fidalgo, J.F. (2022) Optimal approximate choice designs for a two-step coffee choice, taste and choice again experiment. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 71(5), 1895–1917. Available from: https://doi.org/10.1111/rssc.12601