

Article

Utility in Time Description in Priority Best–Worst Discrete Choice Models: An Empirical Evaluation Using Flynn’s Data

Sasanka Adikari  and Norou Diawara * 

Department of Mathematics and Statistics, Old Dominion University, Norfolk, VA 23529, USA; aadik001@odu.edu

* Correspondence: ndiawara@odu.edu

Abstract: Discrete choice models (DCMs) are applied in many fields and in the statistical modelling of consumer behavior. This paper focuses on a form of choice experiment, best–worst scaling in discrete choice experiments (DCEs), and the transition probability of a choice of a consumer over time. The analysis was conducted by using simulated data (choice pairs) based on data from Flynn’s (2007) ‘Quality of Life Experiment’. Most of the traditional approaches assume the choice alternatives are mutually exclusive over time, which is a questionable assumption. We introduced a new copula-based model (CO-CUB) for the transition probability, which can handle the dependent structure of best–worst choices while applying a very practical constraint. We used a conditional logit model to calculate the utility at consecutive time points and spread it to future time points under dynamic programming. We suggest that the CO-CUB transition probability algorithm is a novel way to analyze and predict choices in future time points by expressing human choice behavior. The numerical results inform decision making, help formulate strategy and learning algorithms under dynamic utility in time for best–worst DCEs.

Keywords: discrete choice models (DCM); best–worst scaling; CO-CUB model; paired model; Bellman equation; utility in time (UiT)



Citation: Adikari, S.; Diawara, N. Utility in Time Description in Priority Best–Worst Discrete Choice Models: An Empirical Evaluation Using Flynn’s Data. *Stats* **2024**, *7*, 185–202. <https://doi.org/10.3390/stats7010012>

Academic Editors: Jürgen Pilz, Noelle I. Samia and Dirk Husmeier

Received: 8 January 2024

Revised: 16 February 2024

Accepted: 18 February 2024

Published: 19 February 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Modelling choice behaviors has always been a very challenging part in statistical research. Best–worst scaling (BWs) choice pairs in model behaviors has been coined by Louviere and Woodworth (1990) [1]. With the addition of the attribute and attribute levels, the logistic distributions are convenient tools from analytic and practical aspects; they rely on assumptions of ordinality in the responses that can only be met under limited situations. Furthermore, the computational issues related to the large choices in the best–worst cases hamper patterns in the preferences. Discrete choice experiments (DCE) are one of popular models used to quantify the inspiration of attributes, which characterize the choice options. In the literature, Train (2009) [2] present multiple models based on different assumptions about the distribution of random components. In some of his suggested models, the error terms seem to be homogeneous and uncorrelated. Lancsar et al. (2013) [3] used a form of choice experiment, “Best–Worst Discrete Choice Experiments”, which is designed in a way by asking respondents to choose not only the best option in a choice set but also the worst option followed by the best from the remaining options and so on until the implied preference ordering of the option is obtained. Marley and Louviere (2005) [4] and Marley et al. (2008) [5] discuss some of the theoretical aspects of attribute-level best–worst experiments and Flynn et al. (2007) [6] provides detailed instructions for the analysis of such experiments. Following Marley et al. (2008) [5], Street and Knox (2012) [7] constructed a probability function called best–worst choice probability which can be used to find the probability of choice from one state to another. Working et al. (2018) [8] used a conditional logit approach to calculate the transition (of conditional choice) probabilities over time assuming that the consumer behavior is a time-dependent process.

The assumption of “making two choices in two different time points is independent process for an individual” cannot be justified in general. As mentioned in Potoglou et al. (2011) [9], the interaction between attributes and choice pairs needs further piloting and investigation. The dependence in the modelling of the choices made over time is not fully included in many research venues. The dependence in the selection is a wide and fast field, but the theory is biased towards the estimation method. The premise of the traditional choice models assumes of mutually exclusive alternatives; that is, if a choice is made, alternatives are not selected anymore. An alternative would be to build patterns of the preferences and their weighted measures. That option opens a new framework for choices, especially those that are made over time. Many factors shape the choices made. We include a priority in the selection over times and build it in the design. Specific association in the choices is highlighted to show the need to develop progressive models where temporal dependence in the choices is integrated, and utility is computed. In that sense, the innovation we propose is built under the copula theory, as it captures the dependence among the choices made over time periods irrespective of the marginal distributions. We utilized a completely new copula structure call ‘CO-CUB’ to calculate transition probabilities in the BWs choices over time.

The paper is organized as follows. In Section 2, DCMs are reviewed. The design for the BW choice pairs and their utilities are also included under a priority setting. The transition probabilities under CO-CUB are described in Section 3. We present the model’s theory overtime sequence to quantify and measure consumer behavior and derive utilities using the Markov decision process (MDPs) in Section 4. Section 5 is about the application. Empirical data from Flynn et al. (2007) [6] are simulated and our methods are tested, which have the ability to estimate selected significant parameters such as the “feeling” and “uncertainty” of human beings when making a choice. The literature on Copula for DCE is at its infancy by utilizing our newly constructed probability model to analyze the choices of individuals, comparisons are made, and we end up with a conclusion.

2. Preliminary Results for Best–Worst Discrete Choice Modelling

Best–worst scaling experiments are modified DCEs to elicit further information about the best and worst product, or best and worst attributes and attribute-levels of a product. Respondents are asked make choices between different attributes within a same product; that is the alternatives are the attributes. In the experiment, each of the alternatives is described by a set of K attributes or characteristics of the product or scenario being modelled. Further, each attribute contains l_k levels. That is, each product is represented by a profile $X = (A_1, A_2, \dots, A_K)$, where A_i represents an attribute level of a choice. Consider K profiles such that $P_1 = (A_{11}, A_{12}, \dots, A_{1K}), P_2 = (A_{21}, A_{22}, \dots, A_{2K}), \dots, P_G = (A_{G1}, A_{G2}, \dots, A_{GK})$. For every profile, the choice set is then, $C_X = [(x_1, x_2), \dots, (x_1, x_K), (x_2, x_3), \dots, (x_{K-1}, x_K), (x_1, x_2), \dots, (x_K, x_{K-1})]$, where the first attribute level is the best and the second is the worst, called best–worst pairs. This model is designed in such way that the choice pair (x_a, x_b) , x_a , and x_b cannot be acquired from the same attribute. From the choice set C_X , the respondent determines a BW choice pair from τ such pairs, where $\tau = K(K - 1)$. If we consider G products, that is, G profiles and G associated choice sets, then we can define a finite set S that is $S = \{C_1, C_2, \dots, C_G\}$ such that $C_i \neq C_j$ for all $i \neq j$. The BW choice pairs are illustrated in Figure 1 and shows all the possible combinations.

Total number of attribute levels

$$L = \sum_{k=1}^K l_k$$

and the total number of unique attribute-level pairs

$$Z = \sum_{k=1}^K l_k(L - l_k), \quad (1)$$

where l_k is the number of levels for k^{th} attribute, where $1 \leq k \leq K$. Equation (1) is also described in Street and Knox (2012) [7].

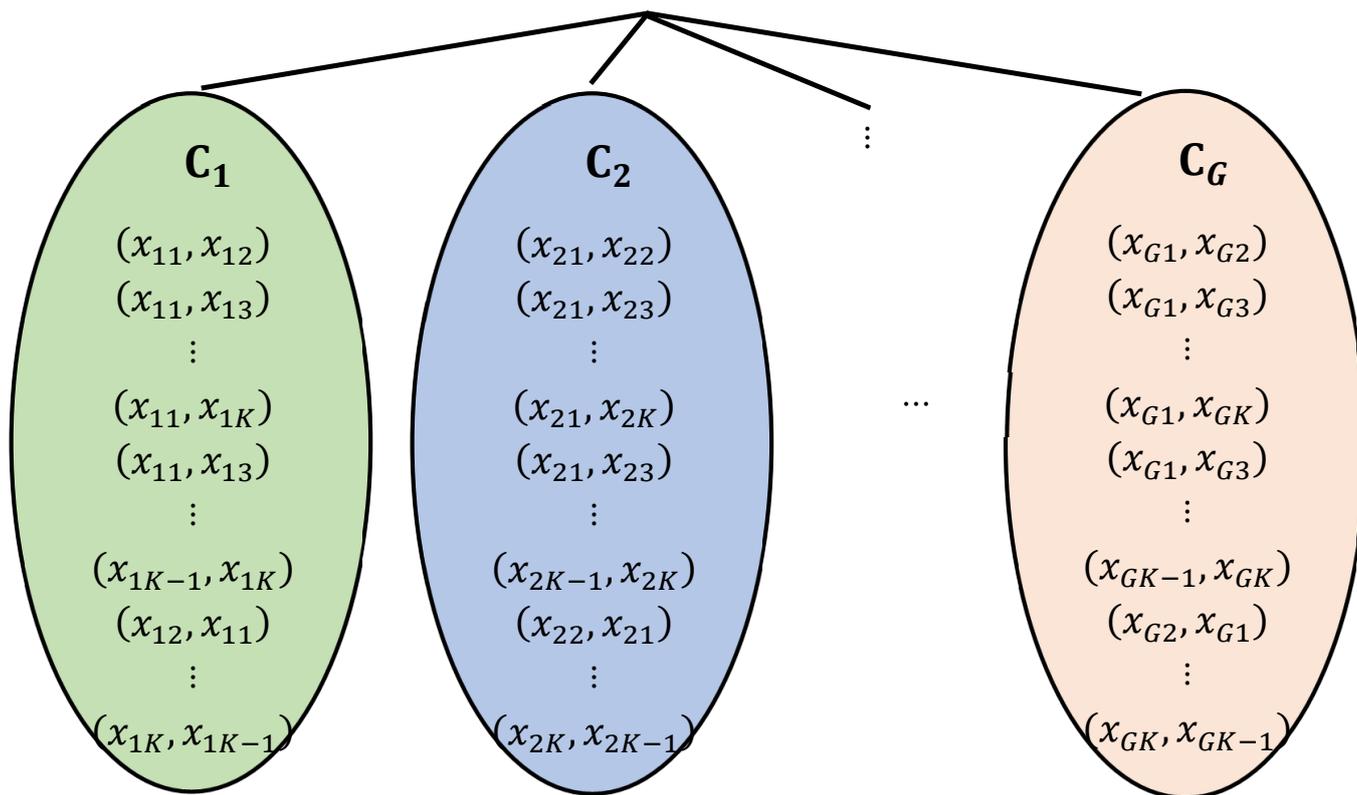


Figure 1. BW Choice pairs in G choice sets, Working et al. (2018) [8].

2.1. Design of Experiment

Assume that the profiles have K attributes, and each attribute has l_k levels, where $k = 1, 2, \dots, K$. The simulation has performed in such a way that each person has to make choices in all sets. The debate concerning optimal designs for DCEs could be enhanced by a closer connection to the established literature on the optimal design of experiments as suggested in Sun et al. (2023) [10]. We used orthogonal design here for the simulation. There are three standard models in best–worst choice model design experiments which are: (1) paired, (2) marginal, and (3) marginal sequential models. In this manuscript, we used the paired model Aizaki and Fogarty (2019) [11] and Marley and Louviere (2005) [4], which assumes that the difference in utility between the two levels represents the greatest utility difference among all τ utility differences.

2.1.1. Paired Model with Attribute and Level Variables

Flynn et al. (2007) [6] used a paired model in his “Quality of life” analysis. In this approach, the attribute variables are created as dummy-coded variables (attribute-specific constants), with reversed signs when the attributes are treated as the worst and the level variables are created as effect coding with a base level for each attribute. The signs of the effect-coded level variables are also reversed when levels are treated as the worst. When estimating this model, an arbitrary attribute variable is omitted, and the coefficient of the omitted attribute variable is normalized to be zero. That is, the coefficients of the remaining attribute variables are estimated relative to the omitted attribute. Unlike dummy coding, effect coding allows the calculation of a coefficient of the base level in each attribute: it is the negative of the sum of the remaining coefficients in each attribute.

Let us consider the general scenario to the formulation for the paired model. Consider the K attributes A_1, A_2, \dots, A_K ; each has l_k levels. Assume that attribute A_K is omitted, and that the last level (l_k) in each attribute (i.e., levels $A_{1l_k}, A_{2l_k}, \dots$ and A_{Kl_k}) is the base level. As in Aizaki and Fogarty (2019) [11], the systematic component of the utility function is then:

$$V = \beta_{A_1}M_{A_1} + \dots + \beta_{A_{K-1}}M_{A_{K-1}} + \beta_{A_{11}}M_{A_{11}} + \dots + \beta_{A_{1(l_k-1)}}M_{A_{1(l_k-1)}} + \dots + \beta_{A_{K(l_k-1)}}M_{A_{K(l_k-1)}}, \tag{2}$$

where the β s are the coefficients to be estimated and Ms are the design matrices associated with the attributes and attribute levels. The coefficient of the attribute variable corresponding to A_K is zero, and the coefficients of the base level $A_{jl_k}, j = 1, 2, \dots, K$ can be calculated by $A_{jl_k} = -\sum_{i=1}^{k-1} \beta_{A_{ji}}$ for $j = 1, 2, \dots, K$. Detailed calculations are shown in the application section. M is composed of indicators for the best and worst attributes and attribute levels. Let M be the $Z \times P$ design matrix, where $Z = \sum_{k=1}^K l_k(L - l_k)$ and $P = Kl_k - 1$. Columns correspond to attribute A_K and the base levels ($A_{1l_k}, A_{2l_k}, \dots$ and A_{Kl_k}) are omitted from the design matrix. Consider the choice pair $(x_{ij}, x_{ij'})$ from the choice set C_i , for $i = 1, 2, \dots, G, j \neq j' = 1, 2, \dots, K$, and $1 \leq x_{ij} \leq l_k$. For

$$M_{A_h} = \begin{cases} 1, & \text{if } x_{ij} \in A_h \text{ for } h = 1, 2, \dots, K - 1 \\ -1, & \text{if } x_{ij'} \in A_h \text{ for } h = 1, 2, \dots, K - 1 \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

Let $M_{A_h x_{ih}}$ be the data for the attribute level $1 \leq x_{ih} \leq l_h, h = 1, 2, \dots, K - 1$ (non-base levels) within attribute $A_j, \forall j = 1, 2, \dots, K$. Referring to the choice pair $(x_{ij}, x_{ij'})$, the corresponding data for the attribute-levels are given by,

$$M_{A_h x_{ih}} = \begin{cases} 1, & \text{if } x_{ij} \in A_h \text{ for } h = 1, 2, \dots, K - 1 \\ -1, & \text{if } x_{ij'} \in A_h \text{ for } h = 1, 2, \dots, K - 1 \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

Within attribute $A_j, \forall j = 1, 2, \dots, K$, the base levels ($A_{1l_k}, A_{2l_k}, \dots$ and A_{Kl_k})

$$M_{A_{1l_k}} = M_{A_{2l_k}} = \dots = M_{A_{Kl_k}} = \begin{cases} -1, & \text{if } x_{ij} \in A_K \\ 1, & \text{if } x_{ij'} \in A_K \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

Table 1 shows the design matrix for the based on data from Flynn et al. (2007) [6] ‘Quality of Life experiment.’ The estimated parameters of the choices are given in Table 2. The BW choice pairs are then arranged according to their popularity (with highest number of picks) in Table 3.

Table 1. Design Matrix based on Equations (3)–(5).

B	W	M_A	M_B	M_C	M_D	M_{A1}	M_{A2}	M_{A3}	M_{B1}	M_{B2}	M_{B3}	M_{C1}	M_{C2}	M_{C3}	M_{D1}	M_{D2}	M_{D3}	M_{E1}	M_{E2}	M_{E3}
A_3	D_2	1	0	0	-1	0	0	1	0	0	0	0	0	0	0	-1	0	0	0	0
A_1	B_3	1	-1	0	0	1	0	0	0	0	-1	0	0	0	0	0	0	0	0	0
E_3	C_3	0	0	-1	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	1
A_2	C_4	1	0	-1	0	0	1	0	0	0	0	1	1	1	0	0	0	0	0	0
D_2	E_3	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	-1
A_1	B_4	1	-1	0	0	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0
A_4	B_2	1	-1	0	0	-1	-1	-1	0	-1	0	0	0	0	0	0	0	0	0	0
A_1	E_4	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1
\vdots																				
E_4	A_4	-1	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	-1	-1	-1

Table 2. Estimated parameters.

Parameters	$\hat{\beta}_{Flynn}$	SE_{Flynn}	$\hat{\beta}_{Sim}$	SE_{Sim}	Parameters	$\hat{\beta}_{Flynn}$	SE_{Flynn}	$\hat{\beta}_{Sim}$	SE_{Sim}
Constant	−0.3067	0.075	*	*	Security All	0.9511	*	0.8199	*
Attachment	0.8105	0.0803	0.8655	0.0575	Enjoyment None	−0.8888	0.1286	−0.8327	0.0759
Security	*	*	*	*	Enjoyment Little	−0.3367	0.1632	−0.3535	0.0773
Enjoyment	0.2632	0.101	0.2684	0.0563	Enjoyment Lot	0.6561	0.1493	0.5862	0.0781
Role	0.1908	0.0974	0.2294	0.0565	Enjoyment All	0.5695	*	0.60007	*
Control	0.1076	0.0971	0.1217	0.0564	Role None	−0.8956	0.1239	−0.7697	0.0761
Attachment_None	−1.9678	0.1129	−1.9633	0.0775	Role Few	−0.0277	0.1532	0.1086	0.0776
Attachment_Little	0.1694	0.1012	0.1811	0.0756	Role Many	0.4435	0.1363	0.2616	0.0784
Attachment_Lot	0.9053	0.0905	1.0493	0.0811	Role All	0.4798	*	0.3995	
Attachment_All	0.8932	*	0.7329	*	Control None	−0.8085	0.1122	−0.7149	0.0763
Security_None	−0.6123	0.118	−0.6953	0.0764	Control Few	0.0835	0.1596	0.1505	0.0774
Security_Little	−0.3761	0.1302	−0.2449	0.077	Control Many	0.278	0.1376	0.2854	0.0785
Security_Lot	0.0373	0.1153	0.1203	0.0779	Control All	0.4471	*	0.2791	*

$\hat{\beta}_{Flynn}$ —Estimated parameters from Flynn’s experiment. $\hat{\beta}_{Sim}$ —Estimated parameters from the simulation. The * represent the base levels in our simulated data.

Table 3. Arrangement of BW choice pairs in choice set 1, based on their popularity.

Best Attribute	Worst Attribute	Best Level	Worst Level	No. of Picks	Rank
Attachment	Security	Attach_all	Security_none	2629	1
Attachment	Role	Attach_all	Role_few	1204	2
Enjoyment	Security	Enjoyment_all	Security_none	1075	3
Attachment	Control	Attach_all	Control_many	994	4
Control	Security	Control_many	Security_none	700	5
Attachment	Enjoyment	Attach_all	Enjoyment_all	629	6
Role	Security	Role_few	Security_none	509	7
Enjoyment	Role	Enjoyment_all	Role_few	498	8
Enjoyment	Control	Enjoyment_all	Control_many	400	9
Control	Role	Control_many	Role_few	341	10
Role	Control	Role_few	Control_many	206	11
Control	Enjoyment	Control_many	Enjoyment_all	162	12
Security	Role	Security_none	Role_few	130	13
Role	Enjoyment	Role_few	Enjoyment_all	122	14
Enjoyment	Attachment	Enjoyment_all	Attach_all	117	15
Security	Control	Security_none	Control_many	89	16
Control	Attachment	Control_many	Attach_all	62	17
Security	Enjoyment	Security_none	Enjoyment_all	58	18
Role	Attachment	Role_few	Attach_all	51	19
Security	Attachment	Security_none	Attach_all	24	20

2.2. Utility Function–Attribute–Level Best–Worst Design

The utility function as described in McFadden (1974) [12] for the i^{th} consumer/individual selecting the j^{th} choice is given as:

$$U_{ij} = V_{ij} + \varepsilon_{ij} \quad (6)$$

where U_{ij} is the utility for i^{th} consumer selecting j^{th} choice. V_{ij} captures the systematic component and ε_{ij} captures the error component. The distribution of the error terms is studied in McFadden (1974) [12]. He proposed the type I extreme value distribution (Gumbel distribution) for the error terms, which modelled data from the conditional logit model. However, this utility function can only accommodate single choice cases. Under the random utility theory, the probability of an alternative is based on the utility as defined in Equation (6); Flynn et al. (2007, 2008) [6,13] and Louviere et al. (2008) [14] provided the utility function for best–worst choice models. Here, we consider the choice set $C, (x_{ij}, x_{ij'})$ being the chosen pair, and the utility for choosing this pair within set C is then given by:

$$U_{ijj'} = V_{ijj'} + \varepsilon_{ijj'}, \quad (7)$$

where $U_{ijj'}$ is the utility value for i^{th} consumer selecting j^{th} choice as the best and j'^{th} as the worst. $V_{ijj'}$ is the corresponding systematic component and $\varepsilon_{ijj'}$ is the error term. The systematic component in Equation (7) can be calculated under the above-described paired model (Equation (2)).

3. Transition Probability

In the context of choice models, we say that the choices are not independent of an irrelevant alternative (IIA) and they are also time dependent. Therefore, we may not achieve an accurate overall performance by not accounting for the relationship between the present and future decisions and their outcomes. Our interest is with discrete time finite horizon Markov Decision Processes (MDP); that is, $t = 1, 2, \dots, T$ where T is a fixed number of time periods. The rewards are maximized by the best sequential decisions over time, making MDPs a dynamic optimization tool as used in Blanchet (2016) [15] to identify the right choices of substitution behaviors of consumers. Let $s_t \in S$ be the states occupied at time t , $r_t(s_t)$ which is the reward associated with s_t , and $d_t(r_t, s_t)$ is the decision based on the possible rewards and states at time t . The decision process maps the movement from one state to another over time t based on rewards received and an optimal decision set. As the decision process is Markovian, the transition probability to the next state, $S' = S_{t+1}$ based solely on the decision made at the current state, $S = S_t$ is $P_{SS'} = P[S_{t+1} = S' | S_t = S]$, where $t = 1, \dots, T$. The conditional probability of state transition from S to S' is defined as:

$$P(s_{t+1}|s_t) = P(s_{t+1} = s' | s_t = s) = P_{ss'} \quad (8)$$

The transition probability of state s' is not homogeneous. As described in Equation (8), it depends on the action taken in the state S . Factors such as ‘feeling’ and ‘uncertainty’ of an individual can be highly affected to this probability. Further, the human behavior has a tendency to select options that have immediate gratification or those with the nearest range rather than selecting alternatives further away in time. Those are some of challenging constraints in terms of building a transition probability function. Further, the dependent structure of adjacent time probabilities needs to be handled carefully. Based on McFadden (1974) [12], the conditional logit model on the error term, Working et al. (2018) [7] establish a time-dependent probability formula while avoiding the IIA condition. It is reasonable as an initiation, but the dependent structure of the choice at the next time point or the above-mentioned ‘close at hand’ constraint cannot be seen in that transition matrix. In this manuscript, we introduced copula-based distribution with a CUB marginal, which is a

combination of a discrete uniform and shifted binomial distributions introduced by Piccolo (2003) [16]. This model is a totally new approach for discrete choice modelling.

3.1. Copula Methods

Many stochastic models define the relation between X and Y through expectations, $E(Y|X = x) = \alpha + \beta x + \varepsilon$. Copulas are functions that connect multivariate distributions to their one-dimensional margins. If F is an m -dimensional cumulative distribution function with one-dimensional margins, F_1, \dots, F_d , then there exists an m -dimensional copula C such that $F(y_1, \dots, y_d) = C(F_1(y_1), \dots, F_d(y_d))$. The case $d = 2$ has attracted special attention; Trivedi and Zimmer (2005) [17]. Copula modelling demonstrates that practical implementation and estimation is relatively straightforward despite the complexity of its theoretical foundations. The copula approach for constructing joint distributions has gained popularity in recent years in applied research in finance, civil engineering, medicine, climate, and weather research and many more. Copula models have been widely used for modelling and characterizing dependence structure in multivariate data. The extension to handle discrete and continuous forms of data offers contingency to model the probability of best–worst pair selection probability.

3.2. CUB Probability Model

Piccolo (2003) [16], Piccolo and Simone (2019) [18], and D’Elia and Piccolo (2005) [19] introduce a radical model which is a Combination of a discrete Uniform and a shifted Binomial distribution (CUB). CUB jointly considers two latent components, called ‘feeling’ and ‘uncertainty’. General references for the statistical background for CUB are included in Iannario (2012) [20] and Iannario and Piccolo (2010) [21]. The stingy parameterization and the ease of estimation and interpretation make CUB models a very useful tool for choice data analyses. Piccolo and Simone (2019) [18] asserts that there is a possibility to fit statistical models to ordinal/choice data without the need of explanatory covariates. Our interest is to use CUB marginals for copula distribution to construct a transition probability matrix which can be used to identify human choice behavior in future time points, given the present state of choice. An interesting attempt at defining a bivariate CUB distribution using a multivariate model with fixed margins is made by Andreis and Ferrari (2013) [22]. The CUB is a class of mixture models, possibly involving covariates, developed as a new approach for modelling discrete choice processes. CUB models consider two components, which are the ‘uncertainty’ and ‘feeling’ of humans when making a choice. The intrinsic ‘uncertainty’ in choosing an item is modelled through a discrete uniform variable. The latent process leading to the choice is directed by the subjective ‘feeling’ and modelled using a shifted binomial distribution. The probability of observing a particular response $r = 1, 2, \dots, m$, with m known. Assuming $m > 3$ to ensure identifiability, the mixture probability model is described in Piccolo (2003) [16] and D’Elia and Piccolo (2005) [19]. It is given as:

$$P(R = r|\pi, \xi) = \pi \binom{m-1}{r-1} \xi^{m-r} (1-\xi)^{r-1} + (1-\pi) \frac{1}{m}; \quad r = 1, 2, \dots, m, \quad (9)$$

with $\pi \in (0; 1]$ and $\xi \in (0; 1]$.

Here, π defines the weights of discrete uniform shifted binomial and it is inversely related to the amount of uncertainty in the choice decisions. The ξ is related to personal preferences and measures the strength of feeling. In CUB models, there is non-uniqueness of the copula representation and the non-uniqueness stems from the fact that marginal distribution functions, which are not strictly monotonically increasing, rather monotonically non-decreasing, do not possess an inverse in the usual sense, rather a pseudo-inverse; Nelsen (2007) [23], Genest et al. (2007) [24]. Because of that reason, we must be precisely cautioned when handling copulas with non-continuous margins, which is in the CUB model.

3.3. The CO-CUB Model (Copula-Based CUB)

Andreis and Ferrari (2013) [22] defines a multidimensional extension of CUB in Equation (9), called CO-CUB model, as a multivariate copula with discrete margins, each following a CUB distribution. A k -dimensional ($k \geq 2$) CO-CUB model with copula C is a multivariate discrete variable with margins $R_i \sim CUB(\pi_i, \zeta_i)$, $i = 1, \dots, k$, each with support $\{1, \dots, m_i\}$, $m_i > 3$. Joint distribution function is given by:

$$\Psi(r_1, \dots, r_k, \underline{\pi}, \underline{\zeta}, \underline{\alpha}) = P(R_1 \leq r_1, \dots, R_k \leq r_k, \underline{\pi}, \underline{\zeta}, \underline{\alpha}) = C_{\underline{\alpha}}[F_1(r_1; \pi_1, \zeta_1), \dots, F_k(r_k; \pi_k, \zeta_k)], \tag{10}$$

where $\underline{\pi} = (\pi_1, \dots, \pi_k)'$ and $\underline{\zeta} = (\zeta_1, \dots, \zeta_k)'$ as in Equation (6), for a particular choice of copula C , characterized by a parameter $\underline{\alpha} = (\alpha_1, \dots, \alpha_d)'$ taking values in some real d -dimensional space α defining the dependence structure of its components. $F_i(r_i; \pi_i, \zeta_i)$ stands for the distribution function of the i^{th} margin; that is, $F_i(r_i) = P(R_i \leq r_i)$ and the support of the CO-CUB variable is the grid $\{1, \dots, m_1\} \times \dots \times \{1, \dots, m_k\}$. The whole parameter set for a k -dimensional CO-CUB is, then, the ordered triplet $(\underline{\pi}, \underline{\zeta}, \underline{\alpha}) \in [0, 1]^k \times [0, 1]^k \times \alpha$, having the following interpretation: by the definition of copula, being margins of CUB, parameters $(\underline{\pi}, \underline{\zeta})$ hold the same interpretation as in the unidimensional case, while for what concerns the copula parameter $\underline{\alpha}$, its interpretation as a dependence measure relates to the specific copula C adopted in Equation (10). Andreis and Ferrari (2013) [22] justified that the Plackett copula distribution is one of the better candidates for the CO-CUB model.

3.4. The CO-CUB Model with Plackett Copula without Covariates

The Plackett copula is defined as:

$$C_{\alpha}(u_{\pi_1, \zeta_1}, v_{\pi_2, \zeta_2}) = \frac{A_{\alpha}(u_{\pi_1, \zeta_1}, v_{\pi_2, \zeta_2}) - \sqrt{A_{\alpha}^2(u_{\pi_1, \zeta_1}, v_{\pi_2, \zeta_2}) - 4\alpha u_{\pi_1, \zeta_1} v_{\pi_2, \zeta_2} (\alpha - 1)}}{2(\alpha - 1)}, \tag{11}$$

for $\alpha \in (0, 1) \cup (1, +\infty)$,

Here, $u_{\pi_1, \zeta_1} = F_1(r_1) = \sum_{i=1}^{r_1} \pi_1 \left\{ \binom{m_1 - 1}{i - 1} \zeta^{m_1 - i} (1 - \zeta)^{i - 1} + (1 - \pi_1) \frac{1}{m_1} \right\}$, where m_1 is the (fixed) number of categories of the first CUB model. This is similar for v_{π_2, ζ_2} and we will from now on assume $m_1 = m_2 = m > 3$. The copula probability mass function can be obtained as $c_{\alpha}(r_1, r_2) = C_{\alpha}(b_1, b_2) - C_{\alpha}(a_1, b_2) - C_{\alpha}(b_1, a_2) + C_{\alpha}$, where $a_i = F_i(r_i)$ and $b_i = F_i(r_i - 1)$ for $i = 1, 2$. Model parameters are then $\pi_1, \zeta_1, \pi_2, \zeta_2$ and α . Estimation can be performed by inference for the margin method (IFM), as shown in Joe and Xu (1996) [25].

3.5. Setting up the Transition Probability Matrix

Let $F(x, y) = C(x, y)$ be the cumulative function and $\frac{f_{XY}(x, y)}{f_X(x)f_Y(y)} = c(u, v)$, where $f_{Y=j|X=i} = \frac{f(y=j, x=i)}{f(x=i)}$. Then, $f_{Y|X=i} = c(u, v)f_Y(y = j)$. For the notation convenient, let the choice state at time point $t + 1$, $s_{i(t+1)} := s'_i$ and the state at t , $s_{i(t)} := s_i$, where s'_i, s_i are ordered labels (ranks) of the BW choice pairs and $s'_i, s_i = 1, 2, \dots, \tau$; $i = 1, 2, \dots, G$, and time $t = 1, 2, \dots, T$. The transition probability is denoted as $P_{iss'}^t = P^t(s'_i | s_i)$. Since we do not have publicly available data from Flynn et al. (2007) [6] experiment, we have to simulate data under the paired model and distributed them among different choice sets. We simulated a number of different datasets under the same set of estimated regression parameters to prioritize choice pairs based on their popularity. That is, for a given choice set C , we counted the number of times an individual picked the choice pair $(x_{ij}, x_{ij'})$. Finally, all the choices are arranged according to their popularity; that is, from the highest count to lowest count. This proceeding gave elements in a choice set, an ordinal power and it is a necessary requirement to incorporate the CUB model for choice data. Further, we adjusted the function for the repetition of selections in order to keep the transition

probability formula as a probability distribution function. Then, the transition probability formula can be written as

$$c^*(u = g, v = h) = \begin{cases} \frac{c(u=g, v=h) n}{C_p} ; C_p \neq 0 \\ 0 ; C_p = 0 \end{cases} \tag{12}$$

where n is the number of repetitions for a selection of the pair $(x_{ij}, x_{ij'})$ in the selected choice set and $C_p = \sum_{r=1}^N c(u = g, v) \cdot I(c(u = g, v))$, c is the density function of the Plackett copula as shown in Equation (11). Let N be the size of the random sample, $g, h = 1, 2, \dots, \tau$, and τ be the number of BW choice pairs considered in each choice set. Then, the probability transition matrix for the i^{th} choice set is in the form of

$$c_i^*(u = g, v = h) = \begin{pmatrix} P_{i11} & P_{i12} & \dots & P_{i1\tau} \\ P_{i21} & P_{i22} & \dots & P_{i2\tau} \\ \vdots & \vdots & \ddots & \vdots \\ P_{i\tau 1} & P_{i\tau 2} & \dots & P_{i\tau \tau} \end{pmatrix} = (P_{iss'}) \tag{13}$$

$i = 1, 2, \dots, G_i, s, s' = 1, 2, \dots, \tau = K(K - 1).$

$P_{iss'}$ is denoted as the transition probability when a consumer changing his choice selection state s at time t to s' at time $t + 1$. For all the transitions, $\sum_{s'=1}^{\tau} P_{iss'} = 1$ (e.g., the sum of the row values in Equation (13) is 1) and if there are no possible transitions $\sum_{s'=1}^{\tau} P_{iss'} = 0$. The ordering scheme is such that for $|s'_i - s_i| \leq \omega, \sum P_{ss'} > \sum P_{ss''}$ for any other $s'' \neq s'$ for the i^{th} choice set, where ω is some small-time epoch number. That is, in simple words the row sum of the probability of the region $|s'_i - s_i| \leq \omega$ is the highest portion than any probability sum in the same row of the probability transition matrix in the i^{th} choice set. We introduce a constraint $|s'_i - s_i| \leq \omega$ for our copula model because the nature of human behavior is that they have a greater tendency to grab options that are close at hand or those within the nearest range rather than picking up alternatives further away in time. The data were simulated based on that choice behavior assumption; that is, $|s'_i - s_i| \leq \omega$. The transition matrix may be either stationary or dynamic in nature. In this manuscript, we considered only the stationary case. It is important to state that the transition probability matrix is calculated by considering all N individuals in the experiment. The row sum of the transition matrix is kept as 0 if there is no transition from s to s' . These proprieties can be seen in Tables 4 and 5.

4. Utility in Time with Dynamic Programming

In this part, we incorporate MDPs and Dynamic Programming to calculate utility in time with the help of transition probabilities. Bellman (1954, 1956) [26,27] introduced a dynamic programming technique to evaluate the utility value function, also known as Bellman’s equation. At each time step, dynamic programming uses numerical methods to evaluate the value function moving backwards in time. Rust (1994, 2008) [28,29] and Ellickson (2011) [30] presented a numerical method with steps for evaluating DCEs as MDPs.

The value function for the DCEs is defined by at each time t as:

$$V_t(X_t, \varepsilon_t) = \max_{d_t \in D} \sum_{t'=t}^T P_{ss'}^{t'} \left(\gamma^{t'-t} U(X_{t'}, d_{t'}) + \varepsilon(d_{t'}) \right) \Big| X_t, \tag{14}$$

with $P_{ss'} = P(s_{t+1}|s_t) = P(s_{t+1} = s' | s_t = s)$. This may also be defined in terms of the transition in decisions as $P(d_{t+1}|d_t, x_t)$, where $d_t \in D$ is the decision at time t , $U(x_t, d_t)$ is the derived expected utility, and ε is the associated error term at time t , where $t = 1, 2, \dots, T$. The discount utility rate in Equation (14) is given by $\gamma \in (0, 1)$; see Feinberg and Schwartz (1994) [31]. There exist τ value functions for each of the τ alternatives in the experiments evaluated at each time point $t = 1, 2, \dots, T$. The sum goes from t up to T because it is evaluated using a backwards recursive method; that is, we start at the last time point and work our way backwards to earlier time points.

$$\begin{aligned} V_T(X_T) &= U(X_T, d_T) \\ V_{T-1}(X_{T-1}, d_{T-1}) &= U(X_{T-1}, d_{T-1}) + \sum_{d_t \in D} \gamma V_T(X_T) P(d_T | d_{T-1}), \\ V_{T-2}(X_{T-2}, d_{T-2}) &= U(X_{T-2}, d_{T-2}) + \sum_{d_t \in D} \gamma V_{T-1}(X_{T-1}) P(d_{T-1} | d_{T-2}), \\ &\vdots \\ V_t(X_t, d_t) &= U(X_t, d_t) + \sum_{t'=t+1}^T \sum_{d_{t'} \in D} \gamma V_{t'}(X_{t'}) P(d_{t'} | d_t), \end{aligned} \tag{15}$$

where U_t is utility in time (UiT) and V_t is the value function. Based on Working et al. (2018) [8], the initial utility for the selected choice pair can be calculated using:

$$V_T = \left((\hat{\beta}_A + \hat{\beta}_{A_k}) - (\hat{\beta}_{A'} + \hat{\beta}_{A'_k}) \right) \equiv U(X_T, d_T), \tag{16}$$

where $\hat{\beta}_A, \hat{\beta}_{A_k}$ are estimated parameters for best attribute and attribute level and $\hat{\beta}_{A'}, \hat{\beta}_{A'_k}$ are the estimated parameters for worst attribute and attribute level, respectively. Instead of this simple model, in this manuscript we used the paired model which is one of the more appropriate models for the paired data. In fact, to distribute the utility over time for the future time points, further adjustment is needed for the value function. We adjusted the value function in Equation (12) for the time in such a way that

$$V_T = (M_+ \hat{\beta}) \theta_b^t + (M_- \hat{\beta}) \theta_w^t \equiv U(X_T, d_T), \text{ where } M = M_+ + M_- \tag{17}$$

M is the design matrix under the paired model which contains +1, -1, and 0 described in Equations (3)–(5). We broke M into M_+ and M_- , which contains +1s and -1s. Further, we introduce arbitrary time-sensitive constants θ_b^t, θ_w^t to make the initial systematic component vary with time.

5. Application and the Design of the Experiment

A pilot best–worst study was conducted in summer 2005 among people aged 65 and over with the aim of informing a larger quality of life valuation exercise by Flynn et al. (2007) [6]. In this survey, there are $K = 5$ attributes, which are Attachment, Security, Role, Enjoyment, and Control. Each attribute with $l_k = 4$ levels. That is

Attachment: Attach_none, Attach_little, Attach_lot, Attach_all

Security: Security_none, Security_little, Security_lot, Security_all
 Role: Enjoyment_none, Enjoyment_little, Enjoyment_lot, Enjoyment_all
 Enjoyment: Role_none, Role_few, Role_many, Role_all
 Control: Control_none, Control_few, Control_many, Control_all

We simulated a new data set for 100 individuals and estimated regression parameters by using a conditional logit model with the help of R package ‘support.BWS2’; see Aizaki and Fogarty (2019) [11]. The application is from a balance design where all five attributes have four- levels each; there are 320 unique BW choice pairs. We simulated data in such a way that 16 choice sets were used to cage all 320 unique BW choice pairs (see Equation (1)), by keeping each choice set with 20 choice pairs. Each person of the 100 sample should make a choice in each choice set. Figure 2 shows the orthogonal design we used to simulate the data. The shape of the questionnaire design based on the orthogonal array is shown in Figure 3. Such a questionnaire could be used in the case of a real-life data collection process. Table 1 shows the design matrix for the paired model. The attribute ‘Control’ is omitted and the fourth level in each attribute (‘Attach_all’, ‘Security_all’, ‘Enjoyment_all’, ‘Role_all’ and ‘Control_all’) is considered to be the base level.

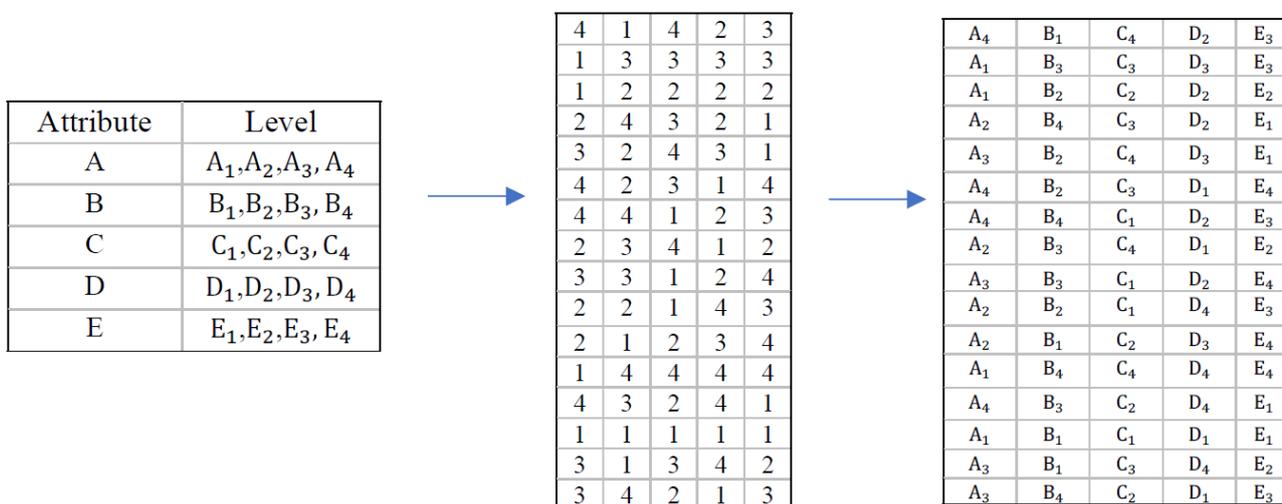


Figure 2. Orthogonal design for Flynn’s experiment flow chart of the 5 attributes, each with 4 levels, with 16 choice sets: e.g., A₄ is the 1st attribute at the 4th level.

Q1			Q2			...	Q16		
Best	Attribute	Worst	Best	Attribute	Worst	...	Best	Attribute	Worst
[]	Attach_all	[]	[]	Attach_none	[]	...	[]	Attach_lot	[]
[]	Security_none	[]	[]	Security_lot	[]	...	[]	Security_all	[]
[]	Enjoyment_all	[]	[]	Enjoyment_lot	[]	...	[]	Enjoyment_little	[]
[]	Role_few	[]	[]	Role_many	[]	...	[]	Role_none	[]
[]	Control_many	[]	[]	Control_many	[]	...	[]	Control_many	[]

Figure 3. Questionnaire design based on the Orthogonal Array.

The parameter estimated by Flynn et al. (2007) [6] and conditional logit regression estimates for the simulation are presented in Table 2. Based on Equation (17), systematic components for each choice and then the corresponding utility can be calculated. Estimations for base levels are calculated by $A_{j|k} = -\sum_{i=1}^{k-1} \beta_{A_{ji}}$ as mentioned in Section 2.1.1. As per the iterative dynamic programming method introduced in Equation (15), with initial utility as described in Equation (16), utility values for a consecutive 10 time periods are calculated recursively by utilizing the transition probabilities and a discount factor $\gamma = 0.95$. As a demonstration, the results for choice set 1 is presented in this manuscript.

The BW choice pairs are arranged in order, as mentioned in Section 3.4. We ran the choice selections algorithm 10,000 times within each choice set to get a proper ordering of elements. Table 3 shows the arrangement for the BW pairs for choice set 1. This arrangement is organized according to the popularity (highest number of picks) of the BW choice pairs. That information is then used to calculate the transition probabilities at the next time point. The probability transition matrix can be calculated utilizing our new model as derived in Equation (12). It is a 20×20 matrix since each choice sets contain 20 choice pairs. Figure 4 shows the highest probability corresponding to each BW choice pair in choice set 1 for three different samples sizes, which $n = 100$, $n = 500$, and $n = 1000$; that is, the highest probable transition slot at time t to $t + 1$. Tables 4–6 show the corresponding transition probability matrices for choice set 1 for sample sizes 100, 500, and 1000, respectively. The constraint region, $|s_i' - s_i| \leq 3$, is highlighted in the probability transition matrix and the highest probability is squared. Nobody picked choice pairs 11, 12, 15, 19, and 20 at time t or $t + 1$ when $n = 100$. That has happened only to choice pair 19 in the $n = 500$ scenario. However, for $n = 1000$ all the choice pairs have been selected. Table 7 shows the utilities (UiT) calculated for time points $t = 1, 2, \dots, 10$. Figure 5 shows the utility–time distribution of selected choices pairs (2, 7, 14, 16, and 18) in choice set 1. It highlights the dynamic utility in time.

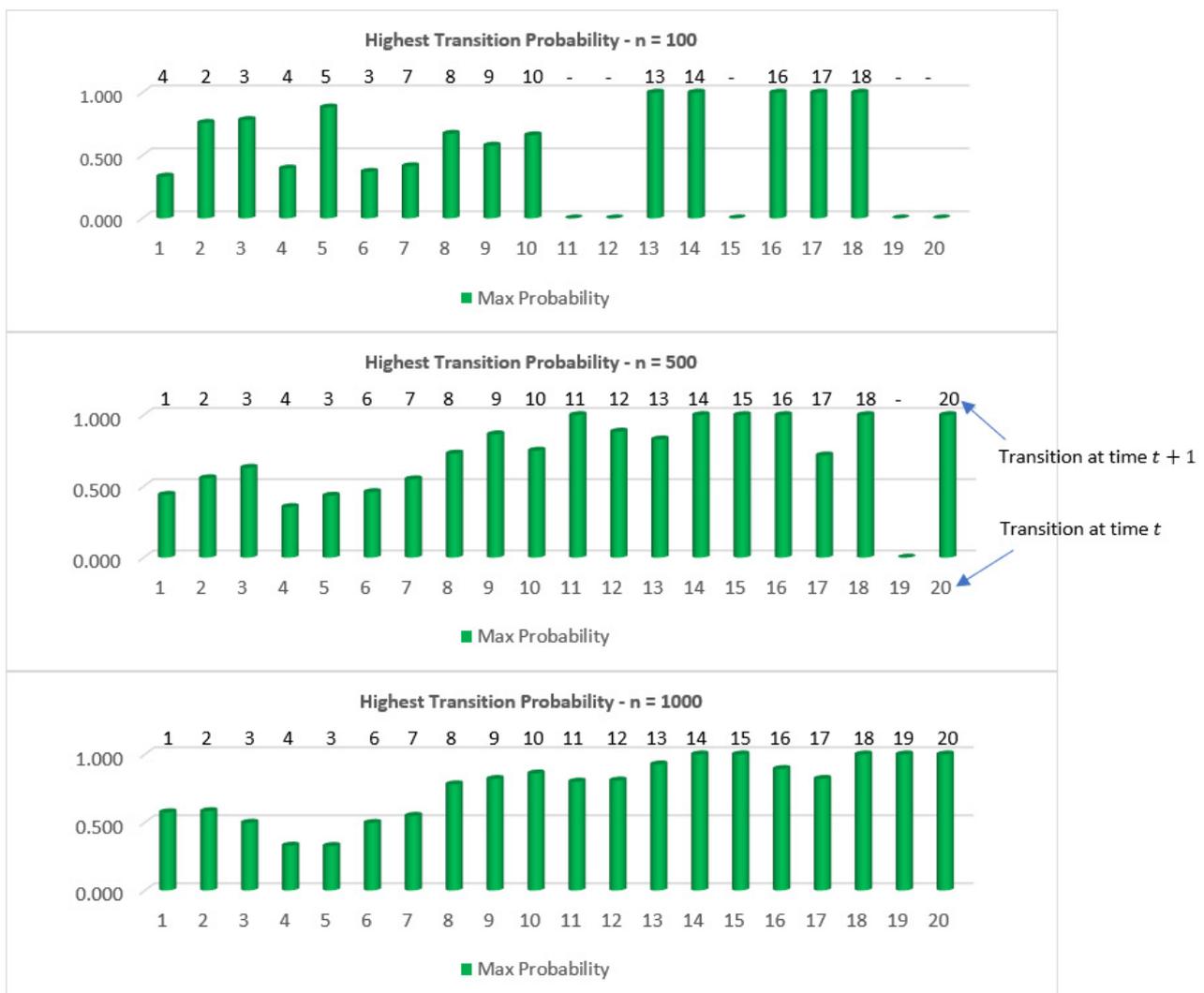


Figure 4. Most probable transition slots from time t to $t + 1$.

Table 7. Expected utilities in time (UiT) for choice set 1.

Choice	t = 1	t = 2	t = 3	t = 4	t = 5	t = 6	t = 7	t = 8	t = 9	t = 10
1	1.2346	1.2613	1.3110	1.3983	1.5419	1.7660	2.1003	2.5814	3.2520	4.1576
2	0.9365	1.0047	1.1107	1.2680	1.4947	1.8143	2.2575	2.8627	3.6758	4.7452
3	0.8028	0.7827	0.7642	0.7572	0.7739	0.8287	0.9381	1.1201	1.3917	1.7638
4	1.2050	1.2500	1.3260	1.4446	1.6215	1.8769	2.2366	2.7325	3.4027	4.2880
5	0.5933	0.6260	0.6770	0.7530	0.8628	1.0178	1.2323	1.5241	1.9134	2.4200
6	0.9295	0.9248	0.9264	0.9418	0.9804	1.0536	1.1745	1.3574	1.6166	1.9616
7	0.9481	0.9799	1.0310	1.1087	1.2231	1.3867	1.6152	1.9276	2.3447	2.8856
8	1.0941	1.1007	1.1168	1.1484	1.2032	1.2902	1.4205	1.6063	1.8600	2.1914
9	1.1719	1.1818	1.2027	1.2404	1.3022	1.3971	1.5357	1.7300	1.9920	2.3310
10	0.7217	0.7911	0.8881	1.0144	1.1734	1.3711	1.6166	1.9242	2.3144	2.8145
11	0.3888	0.4186	0.4630	0.5224	0.5973	0.6890	0.8003	0.9358	1.1030	1.3135
12	0.1282	0.1306	0.1330	0.1335	0.1301	0.1200	0.0997	0.0648	0.0096	-0.0736
13	0.4320	0.4390	0.4430	0.4389	0.4205	0.3802	0.3089	0.1959	0.0295	-0.2022
14	0.7528	0.7431	0.7243	0.6911	0.6365	0.5516	0.4256	0.2452	-0.0044	-0.3383
15	0.3652	0.3564	0.3409	0.3160	0.2779	0.2214	0.1397	0.0234	-0.1402	-0.3686
16	1.4454	1.4596	1.4725	1.4786	1.4711	1.4417	1.3807	1.2765	1.1157	0.8841
17	0.8189	0.8620	0.9079	0.9462	0.9642	0.9469	0.8772	0.7351	0.4989	0.1469
18	1.5412	1.4548	1.3269	1.1446	0.8907	0.5427	0.0715	-0.5588	-1.3894	-2.4609
19	1.0774	1.0947	1.1169	1.1398	1.1576	1.1636	1.1492	1.1046	1.0170	0.8699
20	1.2778	1.2731	1.2608	1.2341	1.1843	1.0997	0.9652	0.7611	0.4613	0.0295

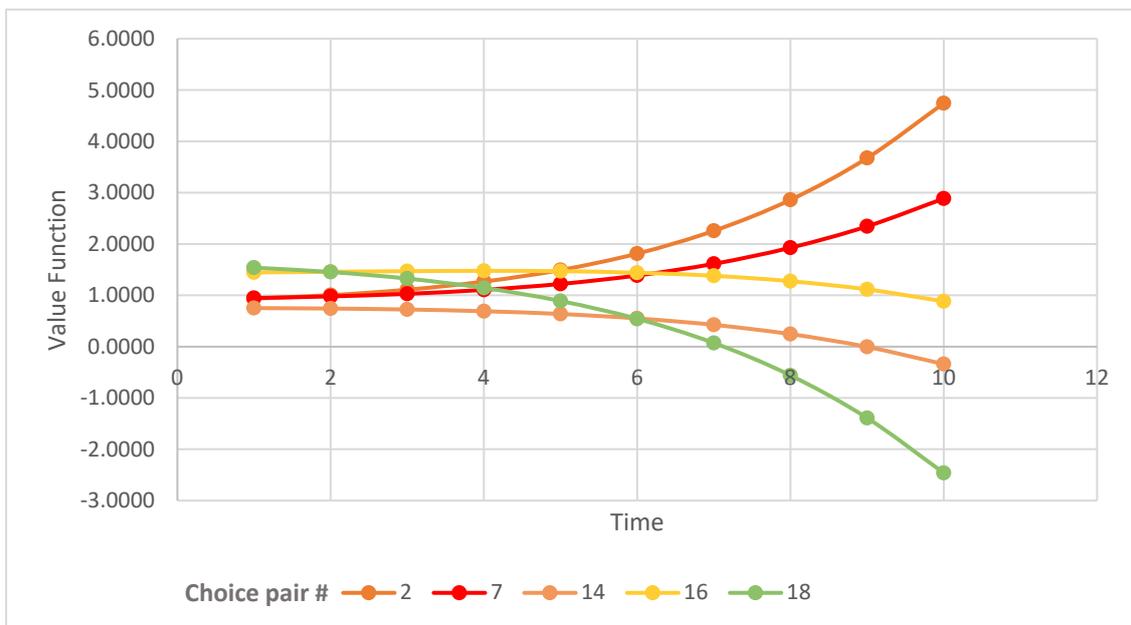


Figure 5. Utility–Time graph for some selected BW choice pairs in choice set 1.

6. Conclusions

In general, the choices are made at different time points and are not IIA for any individual. They depend on not only qualitative factors such as priority setting, feeling, and

uncertainty, but also quantitative covariates such as measure of sequential selection probability in the process. By considering a specific popular model, we capture and monitor the dependent structure and calculate the transition probability between consecutive choices at adjacent time points. We were able to include parameters like feeling and uncertainty to the model. This new approach to calculating transition probabilities extend the ideas proposed by Piccolo (2003) [16] and Piccolo et al. (2019) [32]. Further, we could justify our initial assumption of the human choice behavior of $|s'_i - s_i| \leq \omega$, that they have a greater tendency to grab options that are close at hand or those within the nearest range rather than picking up alternatives further away in time.

Data to quantify the BW in the DCE models are mainly found outside of academia. In this paper, we have simulated the data. Such an issue is one limitation of the paper, but the approach has the advantage of bringing and completing the ranking. The paper could also gain if more covariates were given. Based on Flynn's 'Quality of life experiment', we choose value ω as 3. The accuracy of the CO-CUB model can be further improved relaxing that constraint and by adding more latent covariates and updating knowledge of the ranking labels of the choice pairs within the choice set. We leave such ideas for future research.

Author Contributions: Conceptualization, S.A. and N.D.; methodology, S.A. and N.D.; software, S.A. and N.D.; validation, S.A. and N.D.; formal analysis, S.A. and N.D.; investigation, S.A. and N.D.; data curation, S.A. and N.D.; writing—original draft preparation, S.A. and N.D.; writing—review and editing, S.A. and N.D.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data was simulated.

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

References

1. Louviere, J.J.; Woodworth, G.G. *Best Worst Scaling: A Model for Largest Difference Judgments*; Working Paper; Faculty of Business, University of Alberta: Edmonton, AB, Canada, 1990.
2. Train, K.E. *Discrete Choice Methods with Simulation*; Cambridge University Press: Cambridge, UK, 2009.
3. Lancsar, E.; Louviere, J.; Donaldson, C.; Currie, G.; Burgess, L. Best worst discrete choice experiments in health: Methods and an application. *Soc. Sci. Med.* **2013**, *76*, 74–82. [[CrossRef](#)]
4. Marley, A.A.; Louviere, J.J. Some probabilistic models of best, worst, and best–worst choices. *J. Math. Psychol.* **2005**, *49*, 464–480. [[CrossRef](#)]
5. Marley, A.A.J.; Flynn, T.N.; Louviere, J.J. Probabilistic models of set-dependent and attribute-level best–worst choice. *J. Math. Psychol.* **2008**, *52*, 281–296. [[CrossRef](#)]
6. Flynn, T.N.; Louviere, J.J.; Peters, T.J.; Coast, J. Best–worst scaling: What it can do for health care research and how to do it. *J. Health Econ.* **2007**, *26*, 171–189. [[CrossRef](#)]
7. Street, D.J.; Knox, S.A. Designing for attribute-level best-worst choice experiments. *J. Stat. Theory Pract.* **2012**, *6*, 363–375. [[CrossRef](#)]
8. Working, A.; Alqawba, M.; Diawara, N.; Li, L. Time Dependent Attribute-Level Best Worst Discrete Choice Modelling. *Big Data Inf. Anal.* **2018**, *3*, 55–72. [[CrossRef](#)]
9. Potoglou, D.; Burge, P.; Flynn, T.; Netten, A.; Malley, J.; Forder, J.; Brazier, J.E. Best–worst scaling vs. discrete choice experiments: An empirical comparison using social care data. *Soc. Sci. Med.* **2011**, *72*, 1717–1727. [[CrossRef](#)]
10. Sun, T.; Chen, H.; Gao, Y.; Xiang, Y.; Wang, F.; Ni, Z.; Wang, X.; Huang, X. Best-Worst Scaling Survey of Inpatients' Preferences in Medical Decision-Making Participation in China. *Healthcare* **2023**, *11*, 323. [[CrossRef](#)] [[PubMed](#)]
11. Aizaki, H.; Fogarty, J. An R package and tutorial for case 2 best–worst scaling. *J. Choice Model.* **2019**, *32*, 100171. [[CrossRef](#)]
12. McFadden, D. Conditional logit analysis of qualitative choice behavior. *Front. Econom.* **2005**, *1*, 7–32.
13. Flynn, T.N.; Louviere, J.J.; Peters, T.J.; Coast, J. Estimating preferences for a dermatology consultation using best-worst scaling: Comparison of various methods of analysis. *BMC Med. Res. Methodol.* **2008**, *8*, 76. [[CrossRef](#)] [[PubMed](#)]
14. Louviere, J.J.; Street, D.; Burgess, L.; Wasi, N.; Islam, T.; Marley, A.A. Modeling the choices of individual decision-makers by combining efficient choice experiment designs with extra preference information. *J. Choice Model.* **2008**, *1*, 128–164. [[CrossRef](#)]
15. Blanchet, J.; Gallego, G.; Goyal, V. A Markov chain approximation to choice modeling. *Oper. Res.* **2016**, *64*, 886–905. [[CrossRef](#)]
16. Piccolo, D. On the moments of a mixture of uniform and shifted binomial random variables. *Quad. Di Stat.* **2003**, *5*, 85–104.
17. Trivedi, P.K.; Zimmer, D.M. 2 copulas and dependence. *Found. Trends Econom.* **2005**, *1*, 7–32.
18. Piccolo, D.; Simone, R.; Iannario, M. Cumulative and CUB models for rating data: A comparative analysis. *Int. Stat. Rev.* **2019**, *87*, 207–236. [[CrossRef](#)]

19. D'Elia, A.; Piccolo, D. A mixture model for preferences data analysis. *Comput. Stat. Data Anal.* **2005**, *49*, 917–934. [[CrossRef](#)]
20. Innario, M. CUBE models for interpreting ordered categorical data with overdispersion. *CUBE Models Interpret. Ordered Categ. Data Overdispersion* **2012**, *14*, 137–140.
21. Iannario, M.; Piccolo, D. A new statistical model for the analysis of customer satisfaction. *Qual. Technol. Quant. Manag.* **2010**, *7*, 149–168. [[CrossRef](#)]
22. Andreis, F.; Ferrari, P.A. On a copula model with CUB margins. *Quad. Di Stat.* **2013**, *15*, 33–51.
23. Nelsen, R.B. *An Introduction to Copulas*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2007.
24. Genest, C.; Nešlehová, J. A primer on copulas for count data. *ASTIN Bulletin J. IAA* **2007**, *37*, 475–515. [[CrossRef](#)]
25. Joe, H.; Xu, J.J. The Estimation Method of Inference Functions for Margins for Multivariate Models. 1996. Available online: <https://open.library.ubc.ca/soa/cIRcle/collections/facultyresearchandpublications/52383/items/1.0225985> (accessed on 15 February 2024).
26. Bellman, R. The theory of dynamic programming. *Bull. Am. Math. Soc.* **1954**, *60*, 503–515. [[CrossRef](#)]
27. Bellman, R. Dynamic programming and Lagrange multipliers. *Proc. Natl. Acad. Sci. USA* **1956**, *42*, 767–769. [[CrossRef](#)] [[PubMed](#)]
28. Rust, J. Structural estimation of Markov decision processes. *Handb. Econom.* **1994**, *4*, 3081–3143.
29. Rust, J. Dynamic programming. *New Palgrave Dict. Econ.* **2008**, *1*, 8.
30. Ellickson, P.B.; Misra, S. Structural workshop paper—Estimating discrete games. *Mark. Sci.* **2011**, *30*, 997–1010. [[CrossRef](#)]
31. Feinberg, E.A.; Shwartz, A. Markov decision models with weighted discounted criteria. *Math. Oper. Res.* **1994**, *19*, 152–168. [[CrossRef](#)]
32. Piccolo, D.; Simone, R. The class of CUB models: Statistical foundations, inferential issues and empirical evidence. *Stat. Methods Appl.* **2019**, *28*, 389–435. [[CrossRef](#)]

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.