Portfolio Choices and Cross Effects Designs

James B. Wiley

Victoria University
School of Marketing and International Business
PO Box 600
Wellington
New Zealand
Email: jameswiley@vuw.ac.nz

Harry Timmermans

Eindhoven University of Technology
Urban Planning Group
PO Box 513
5600 MB Eindhoven
The Netherlands
Email: h.j.p.timmermans@bwk.tue.nl

Abstract: This paper argues that certain choice problems in transportation research can be best conceptualized as problems of portfolio choice. It discusses how portfolio choice problems can be studied using stated choice (conjoint) methods. Issues in the construction in the design of experiments, model specification and estimation are discussed. A working example illustrates the basic approach. Finally, possible extensions of the basic approach are identified.

Key words: Portfolio choice, stated preference, experimental design, cross effects

Introduction

Recently, stated preference and choice models found ample application in transportation research. These models involve constructing experimental designs to vary a set of attributes such that the necessary and sufficient conditions to estimate the preference or choice model of interest are satisfied. In the case of stated preference analysis, the goal is to estimate a preference or utility function, implying that respondents are asked to express their degree of preference for the experimentally varied choice alternatives. In case of choice experiments, the goal is also to predict preferences/utilities or to predict market shares of a set of choice alternatives. Stated preference and choice models in transportation research have especially been applied (i) in situations where the choice alternatives were new and by consequence one could not rely on historical, revealed preference, choice data (e,g., a new bridge, toll road, pricing scheme); (ii) in situations where the observed attributes were highly correlated and the use of experimental designs offer the researcher the advantage of controlling between-attributes correlations (e.g., choice of shopping center), and (iii) to derive willingness-to-pay and valuation of time measures used to assess the economic impact of new infrastructure projects. Examples of such applications are [1, 2, 3, 4]. A more detailed review of stated preference and choice models can be found in [5, 6].

Existing applications of stated preference and choice models in transportation research typically estimate utility functions and choice models for choice alternatives described in terms of a set of alternatives. The experimental task is to select from each choice set the most preferred choice alternative. This *single* choice task however does not cover all choice problems relevant to transportation research. One class of choice problems that, to the best of our knowledge, has not received much attention in transportation analysis is portfolio choice. The portfolio choice problem can be defined as the problem of choosing a *combination* of choice alternatives as a function of the attributes of the choice alternatives, and possibly some contextual variables. The composition of the chosen set of alternatives may have a specific effect on the probability of choosing particular combinations of choice alternatives. Examples would be combined destination-transport mode choice, task allocation, composition of activity agendas, and time allocation. These problems have in common that individuals or households have to decide on the combinations of choice alternatives that provide the highest utility, perhaps subject to some constraints. The utility of each choice alternative is not invariant, but in part depends on the presence or absence and attribute values of the other alternatives in a choice set.

In this paper, we will introduce a conjoint-based approach that can be used to analyze and predict portfolio choices. We will first introduce the basics and discuss the key principle underlying this new approach. Next, we will discuss some issues in the design of experiments that need to be adhered to estimate models of portfolio choices. This is followed by a discussion of coding the experimental design and the interpretation of the resulting parameter estimates. We complete the paper with a discussion of some interesting issues of future research.

Conjoint Analysis

In alternative choice based conjoint analysis, a specific combination of alternatives that might describe a choice situation is called a *choice set*. Choice sets are usually presented as abstract representations whose aspects consist solely of specified alternatives. The primary reason for restricting the choice situation in this way is to assure that respondents evaluate each alternative with respect to the same information. Ambiguous and equivocal cues are removed and consequently all respondents have at their disposal the same information, and no more.

The investigator organizes choice alternatives into t systematically constructed choice sets of t_i alternatives, $t_i \leq t$, where t_i is the number of alternatives in set i, $i= 1,2,...,t$. N respondents are presented a series of such sets. Typically, respondents are asked to select the one alternative they consider be best in each set. In many applications, however, it is reasonable for more than one profile to be selected.

Following traditional calculation for random utility choice theory, the probability of $P(A)$, is given by $e^{v_A} /(e^{v_A} + e^{v_0=0})$, where $v_0 = 0$ is the reference point for no choice when the choice set consists of choosing, or not choosing, an alternative, when the choice set consists of choosing, or not choosing, the alternative. The probability of not choosing $\overline{A}$ is $e^{v_0=0} /(e^{v_A} + e^{v_0=0})$. The logit, or log odds of choosing A over no choice is $L[P(A)/P(NC)] = v_A$. For illustration, if $v_A = 0.25$, P(A|A,NC) = 0.56; if $v_A = -.25$, P(A|A,NC) = 0.44.

When the choice set consists of A, B, and no choice: the traditional calculations for "pick-one" choice probabilities are:

$$P(A) = e^{v_A} /(e^{v_A} + e^{v_B} + e^{v_0=0}) \, , \tag{1a}$$

$$P(B) = e^{v_B} /(e^{v_A} + e^{v_B} + e^{v_0=0}) \, , \tag{1b}$$

$$P(NC) = e^{0} /(e^{v_A} + e^{v_B} + e^{v_0=0}). \tag{1c}$$

The respective logits for choosing A and B are: $L[P(A)/P(NC)] = v_A$, $L[P(B)/P(NC)] = v_B$. Note that the utility values for the alternatives are independent of other alternatives in the choice set, i.e., they are "independent from irrelevant alternatives" (IIA).

If $v_A = 0.25$ and $v_B = -0.25$, Equation 1a gives P(A|A,B,NC) = 0.42, Equation 1b gives P(B|A,B,NC) = .25, and Equation 1c gives P(NC|A,B,NC) = 0.33. We compare these results with our portfolio choice results in the next section.

Portfolio Choice

The objective of this paper is to discuss approaches for the design and analysis of portfolio choice experiments. Portfolio choice refers to situations where respondents are given a choice set consisting of t_i alternatives and are allowed to select any combination of them, including none. For example, suppose the choice set consists of two alternatives, A and B. The set of possible choices consists of; A, B; AB, or none of the alternatives.

Choice probabilities for the various combinations of A and B may be calculated as the product of the probabilities of choosing (or not choosing) the component alternatives. For example, the probability of choosing none of the alternatives is $P(\overline{A})*P(\overline{B})$, which is the probability of not choosing A *and* not choosing B in a single alternative choice set. The probability of choosing A equals $P(A)*P(\overline{B})$, where $P(A)$ is the probability of choosing A and $P(\overline{B})$ (B with over-bar) is the probability of choosing not-B, with similar interpretations for other alternatives in a choice set.

For illustration, suppose $v_A = 0.25$ and $v_B = -.25$. According to Equations 1a and 1b the probabilities of A vs. NC are 0.56 vs. 0.44. The probabilities of B vs. NC are 0.44 and 0.56. The probabilities of the respective combinations are:

$$P(A \mid AB) = P(A \mid A, NC)*P(\overline{B} \mid B, NC) = 0.56*0.56 = 0.32. \qquad \text{(2a)}$$

$$P(B \mid AB) = P(\overline{A} \mid A, NC)*P(B \mid B, NC) = 0.44*0.44 = 0.19. \qquad \text{(2b)}$$

$$P(AB \mid AB) = P(A \mid A, NC)*P(B \mid B, NC) = 0.56*0.44 = 0.25. \qquad \text{(2c)}$$

$$P(NC \mid AB) = P(\overline{A} \mid A, NC)*P(\overline{B} \mid B, NC) = 0.44*0.56 = 0.25. \qquad \text{(2d)}$$

Note that the probability of choosing *A* from the choice set {*A*, $\overline{A}$ } (0.56) is equal to the sum of the probabilities for P(*A*) and P(*AB*) given by Equation 2a and 2c, 0.32 + 0.25 = 0.56. The probability of choosing *B* from the choice set {*B*, $\overline{B}$ } (0.44) is equal to the sum of the probabilities for P(B) and P(AB) given by Equation 2a and 2b, 0.19 + 0.25 = 0.44. In other words, the sum of the probabilities for the combinations in which an alternative is present equals the marginal probability of choosing the alternative in the set consisting the alternative and no choice.

The advantage of computing choice probabilities for combinations rather than based on traditional "pick-one" choice is seen by comparing results using Equations 1 and 2. The probability of choosing an alternative, for example A as given by Equation 1a, is not equal to the probability for A as given by Equation 2a. The reason for this is that Equation 1a excludes the logical possibility of choosing the combination of AB from the set consisting of {A, B, NC}. Equations 1 distribute the probability of choosing the combination AB over the three possibilities in Equation set 1, i.e., A, B, or NC. That is, following IIA, the 0.25 percent that goes to the AB combination under Equations 2 is distributed over the three Equation 1 possibilities proportionate to the Equation 2 shares for A, B, and NC. For example, excluding the probability of 0.25 for the AB, the probabilities for A, B, and NC sum to 0.75. The 0.25 share for the AB combination under Equation 2 is allocated under Equation 1 as 0.25*(0.32/0.75) = 0.10. The 0.42 share according to Equation 1a equals the P(A) share calculated using Equation 2a of 0.32 plus the 0.10 share allocated from the excluded AB combination. In general, the probabilities calculated for "pick-one" data using the traditional approach will equal the P(A) probability under the Equation 2 approach plus an allocation of shares of combinations not considered under the "pick-one" formulation.

An Alternative Notation for Portfolio Choice

In what follows, an alternative but equivalent notation simplifies the representation of compound choices. The alternative notation represents the probability of choosing A as $e^{v_A} /(e^{v_A/2} + e^{-v_A/2})$. The probability of choosing not A is $e^{-v_A} /(e^{v_A/2} + e^{-v_A/2})$.

Given the above alternative notation, the probability of choosing the four compound choices from the set t_i = (AB) may be calculated as:

$$P(A) = P(A,\overline{B}) = \frac{e^{v_A/2}}{e^{v_A/2} + e^{-v_A/2}} * \frac{e^{-v_B/2}}{e^{v_B/2} + e^{-v_B/2}}$$

$$P(B) = P(\overline{A},B) = \frac{e^{-v_A/2}}{e^{v_A/2} + e^{-v_A/2}} * \frac{e^{v_B/2}}{e^{v_B/2} + e^{-v_B/2}}$$

$$P(A \& B) = P(A,B) = \frac{e^{v_A/2}}{e^{v_A/2} + e^{-v_A/2}} * \frac{e^{v_B/2}}{e^{v_B/2} + e^{-v_B/2}}$$

$$P(no\,choice) = P(\overline{A},\overline{B}) = \frac{e^{-v_A/2}}{e^{v_A/2} + e^{-v_A/2}} * \frac{e^{-v_B/2}}{e^{v_B/2} + e^{-v_B/2}}$$

In general, there will be 2^{t_i} combinations, where t_i is the number of alternatives in the choice set.

We illustrate the equivalence of the notational approaches for calculating compound choices using the eight combinations that may be formed from three alternatives, A, B, and C. Regardless of notation, the probability of the combination AB may be calculated as $P(A) * P(B) * P(\overline{C})$. The probability using traditional notation is:

$$P(A,B,\overline{C}) = P(AB) = \frac{e^{v_A}}{e^{v_A} + 1} * \frac{e^{v_B}}{e^{v_B} + 1} * (1 - \frac{e^{v_C}}{e^{v_C} + 1}). \qquad (3a)$$

Using the alternative notation of the above paragraph, the product may be represented as:

$$P(A,B,\overline{C}) = P(AB) = \frac{e^{v_A/2}}{e^{v_A/2}+e^{-v_A/2}} * \frac{e^{v_B/2}}{e^{v_B/2}+e^{-v_B/2}} * \frac{e^{-v_C/2}}{e^{v_C/2}+e^{-v_C/2}}. \qquad (3b)$$

The primary advantage of the alternative notation is in the representation of combinations involving the presence and absence of alternatives in a choice set.

Regardless, of notation, the analysis of choice proportions for combinations may be analyzed using traditional logit analysis. For example, the logit of choice proportions for the AB combination to the no choice combination is:

$$\ln(\frac{P(A,B,\overline{C})}{P(\overline{A},\overline{B},\overline{C})} = \ln(\frac{\frac{e^{v_A/2}}{e^{v_A/2}+e^{-v_A/2}} * \frac{e^{v_B/2}}{e^{v_B/2}+e^{-v_B/2}} * \frac{e^{-v_C/2}}{e^{v_C/2}+e^{-v_C/2}}}{\frac{e^{-v_A/2}}{e^{v_A/2}+e^{-v_A/2}} * \frac{e^{-v_B/2}}{e^{v_B/2}+e^{-v_B/2}} * \frac{e^{-v_C/2}}{e^{v_C/2}+e^{-v_C/2}}}) = 2v_A/2 + 2v_B/2 + 0 = v_A + v_B$$

Similarly, the logits to the base of no choice of the remaining six probabilities are:

$$\ln[P(A,\overline{B},\overline{C})/P(no\,choice] = v_A, \ln[P(\overline{A},B,\overline{C})/P(no\,choice] = v_B$$
$$\ln[P(\overline{A},\overline{B},C)/P(no\,choice] = v_C,\ \ln[P(A,\overline{B},C)/P(no\ choice] = v_A + v_C$$
$$\ln[P(\overline{A},B,C)/P(no\ choice] = v_B + v_C$$

Note the straightforward interpretation of the relationship between the values of the alternatives and the logits of combinations.

Table 1 gives the choice proportions for the eight combinations calculated using notations 3a and 3b. Proportions are provided for values of v_A ranging from –4 to +4 while holding v_B = 0.50 and v_C = -0.25. As expected, the alternative notations provide identical values for the probability of picking each combination. Figure 1 illustrates the changes in the proportions of choosing the respective combinations of A, B, and C as the value of A increases. As expected, probabilities of choosing combinations involving alternative A all increase as the value of A increases. Holding the value of the other alternatives constant, probabilities of choosing combinations *not* involving A all decrease as the value of A increases.

Table 1

Probabilities of Choice Combinations Using Traditional and Alternative Notation

Combinations	Value of v_A								
Alternative Notation	-4.00	-3.00	-2.00	-1.00	0.00	1.00	2.00	3.00	4.00
NC	0.21	0.20	0.19	0.16	0.11	0.06	0.03	0.01	0.00
A	0.00	0.01	0.03	0.06	0.11	0.16	0.19	0.20	0.21
B	0.34	0.33	0.31	0.26	0.17	0.09	0.04	0.02	0.01
C	0.16	0.16	0.15	0.12	0.08	0.04	0.02	0.01	0.00
AB	0.01	0.02	0.04	0.09	0.17	0.26	0.31	0.33	0.34
AC	0.00	0.01	0.02	0.04	0.08	0.12	0.15	0.16	0.16
BC	0.27	0.26	0.24	0.20	0.14	0.07	0.03	0.01	0.00
ABC	0.00	0.01	0.03	0.07	0.14	0.20	0.24	0.26	0.27
Tradtional Notation									
NC	0.21	0.20	0.19	0.16	0.11	0.06	0.03	0.01	0.00
A	0.00	0.01	0.03	0.06	0.11	0.16	0.19	0.20	0.21
B	0.34	0.33	0.31	0.26	0.17	0.09	0.04	0.02	0.01
C	0.16	0.16	0.15	0.12	0.08	0.04	0.02	0.01	0.00
AB	0.01	0.02	0.04	0.09	0.17	0.26	0.31	0.33	0.34
AC	0.00	0.01	0.02	0.04	0.08	0.12	0.15	0.16	0.16
BC	0.27	0.26	0.24	0.20	0.14	0.07	0.03	0.01	0.00
ABC	0.00	0.01	0.03	0.07	0.14	0.20	0.24	0.26	0.27

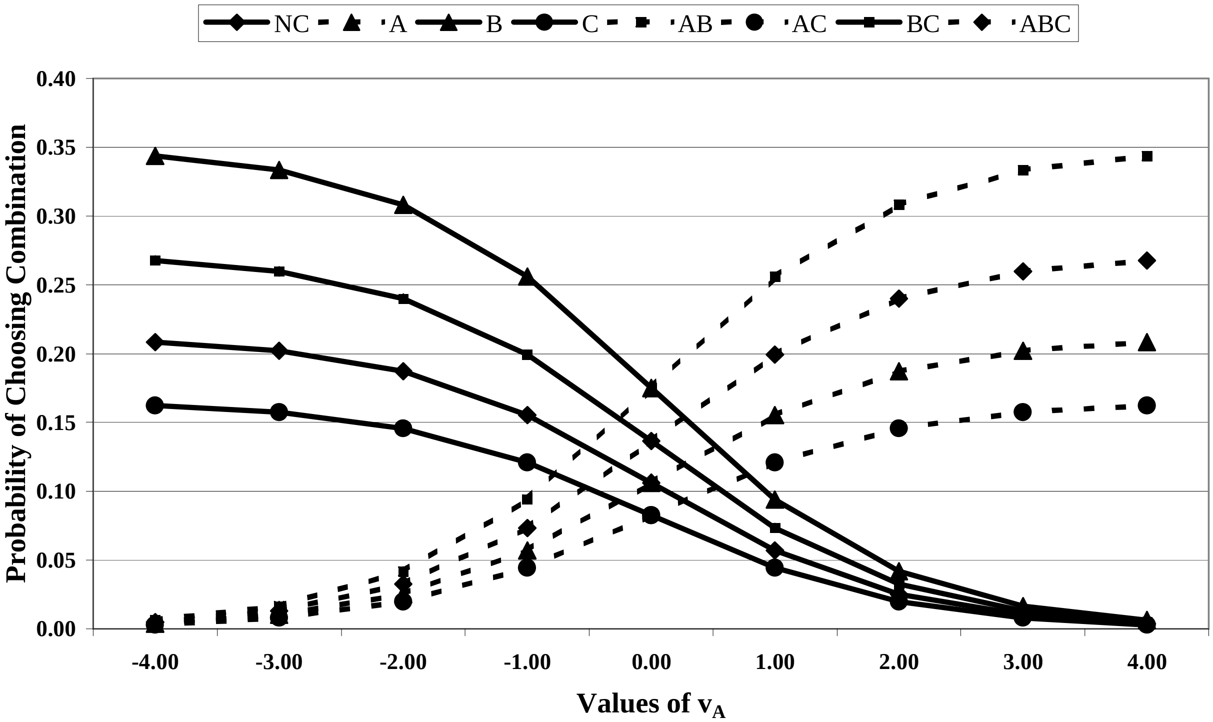
Note: The probabilities of the respective combinations are calculated using forms such as Equation 1 for the traditional notation and Equation 2 for the alternative notation. Results using the respective notations are identical for all values v_A ranging from –4.00 to +4.00. The value for v_B is fixed at 0.50 and the value for v_C is fixed at –0.25.

Impact of Choice Set Composition on Choice

The above results provide the means for analyzing stated preferences for combinations selected from a choice set of size t_i and a given composition. As shown in by the comparison Equation 1 "pick-one" and Equation 2 combinations approach, the Independence-of-Irrelevant-Alternatives (IIA) property applies to the analysis. When collecting stated choices using choice based conjoint measurement, however, respondents are given multiple choices sets of varying size and composition and it is likely that the composition of the set will influence the attractiveness of the alternatives within the set. In order to make the calculation of combinations selected from choice sets having differing composition realistic, it is desirable that the value of alternatives may differ depending on the composition of the choice set in which they appear. We now turn to this problem.

Figure 1

Probabilities of Choosing Combinations of Three Alternatives



Note: Values for v_A (the value of A) range from –4.00 to +4.00. The value for v_B is fixed at 0.50 and the value for v_C is fixed at –0.25. As the value of v_A increases, the probability of choosing all combinations involving A increases and the probabilities of all other combinations decreases.

Cross effects models

Cross effects models have been developed to accommodate the possibility that choice set composition will influence the attractiveness of alternatives in a set. In the simplest, "alternatives only" versions of these models, preference for an alternative is in part an additive function of the other alternatives available in a choice set. To illustrate the "alternatives only" model, imagine a four alternative design, one in which the first two choice sets have the composition {i = D; *k* = 1} and {i = A, C, D; *k* = 2, k = 1, …,number of choice sets}. The cross-effect model for the value a) $V_{D|1}$ (alternative D in set 1) and b) $V_{D|2}$ (alternative D in set 2) may be represented as:

$$v_{i=D,\ j\,|k=1} = -a_{DA} - a_{DB} - a_{DC} + a_{DD} \quad (4a)$$

$$v_{i=D,\ j\,|k=2} = +a_{DA} - a_{DB} + a_{DC} + a_{DD}, \quad (4b)$$

where a_{ij}, $i = j$ are the "own effects" and a_{ij}, $i \neq j$ are the "cross-effects". For example, a_{DA} is the effect that the presence of alternative A in a choice set has on a_{DD}, the "own effect" of D.

With this (1 -1) "effects" coding, the cross-effects in Equation 4 are recognizable as main effects of a 2^{m-1} factorial ANOVA design. The own effect corresponds to a mean. The cross-effects are the main effects of adding the respective other alternatives to a choice set. For example, the parameter $\mathbf{a}_{DA}$ is the differential effect (increase or decrease) that the availability of alternative A in a choice set has on the log-odds of choosing alternative D.

If alternative A offers all the benefits of alternative D, and some additional benefits, then one would expect the availability of alternative A to reduce the likelihood that alternative D would be selected *by more than would occur under IIA* since the consumer could get all the benefits alternative D offers by selecting alternative A. A is a substitute for D in this instance. Here it would be expected that parameter $\mathbf{a}_{DA}$ would be negative. The presence of A in a choice set has the effect of reducing the attractiveness of D.

On the other hand, the option of picking both alternatives in a choice set containing both of them may increase the attractiveness of both. For example, some people who did not choose A in the absence of B, may choose A as part of the AB combination, and vice versa. Alternatives A and B complement one another in this case.

Cross effects may be positive or negative. A positive cross effect occurs when adding an alternative to a set results in a probability of choosing both is *greater* than would be expected if IIA held. A negative cross effect occurs when adding an alternative to a choice set results in a probability of choosing both that is less than would occur if IIA held. Clearly the choice proportions that would prevail if IIA held provide a benchmark that defines equal substitutability between alternatives and cross effects pickup deviations from equal substitutability.

Note that the t_i^2"own" and "cross-effects" of a t_i alternative experiment may be organized as:

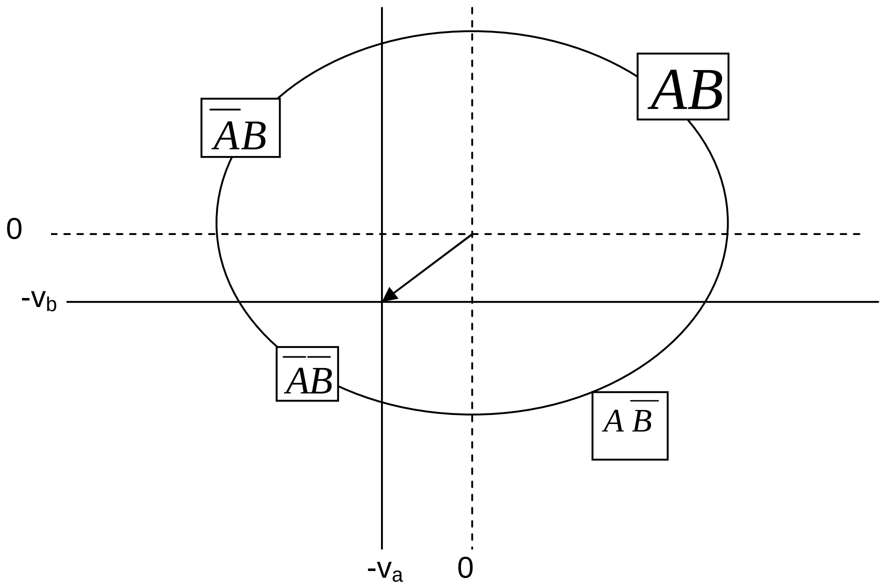
$$V = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1t_i} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2t} \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_{t_i 1} & \alpha_{t_i 2} & \cdots & \alpha_{t_i t_i} \end{bmatrix} . \tag{5}$$

The parameters along the main diagonal are the "own effects." The parameters in the row j (excluding a_{jj}) capture the effects of the availability of other alternatives on alternative j. The parameters in column *j* (excluding a_{jj}) capture the effect that of the availability of alternative *j* on competing alternatives. As noted above, the elements of Eq. 4 may have the same or opposite signs. They are not necessarily symmetric ($a_{ij} = a_{ji}$) or anti-symmetric ($a_{ij} = -a_{ji}$).

For example, in the case of the choice of two alternatives, the four combinations may be represented as shown in Figure 2, which illustrates that the probability of the respective combinations of two alternatives correspond to the quadrature of a bivariate distribution.

Figure 2

Combinations of Choosing Two Alternatives From a Set of Four



The example assumes that the cross-effects for A and B are both positive, v_a and v_b respectively, thus increasing the attractiveness of both alternatives when both are present in a choice set. An equivalent representation to increasing the value of A and B is to view their joint effect as shifting the origin of the distribution in the opposite direction, i.e., $-v_a$ and $-v_b$ units. It is then evident that increasing the attractiveness of *both* A and B increases the probability of choosing P(AB); and reduces the probability of choosing P(A), P(B), or P(NC). On the other hand, a positive value for A and a negative value for B (say v_a and $-v_b$) shifts the origin in a northwesterly direction reducing the probability of P(*AB*) and P(*B*); and increases P(*A*) and P(*NC*). Thus, a variety of observed choice proportions for the respective combination may be accommodated with appropriate origin shift represent by cross effects. The origin shift generalizes. The impact of choice set composition on changes in choice probabilities of three alternatives are modeled by origin shift in a three dimensional distribution, four alternatives by origin shift in a four dimensional distribution, and so forth.

Designing Portfolio Choice Experiments Using Availability Designs

In order to implement the above approach, it must be possible to design experiments in which own and cross effects may be estimated. Availability designs have this capability. Generally, a 2^{t_i} main effect plan and its foldover will provide an availability design. Raghavarao and Wiley [8] used 3-designs to create availability designs. Other work relevant to alternatives only, Availability Designs includes that of Raghavarao, Federer, and Schwager [9] who show that cross effects models are estimable using the 3-designs (doubly balanced incomplete block designs) introduced by Calvin [10]. Bhaumik [11] shows that a 3-design is universally optimal in the class of binary block designs for estimating cross effect models (when random errors are IID with mean zero and variance σ^2). Early applications of these design strategies include Anderson, Borgers, Ettema and Timmermans [12]. A recent paper by Raghavarao and Zhou [13] introduced designs with unequal set sizes called UE 3-designs to estimate cross effects models. They show these designs are universally optimal when random errors are IID with mean zero and variance σ^2.

Anderson and Wiley [7] illustrate that efficient availability designs may be easily generated using Hadamard matrices. A Hadamard matrix **H** is a square matrix of 1's and –1's whose rows and columns are orthogonal, i.e**., H'H = HH' = I.** An important property of Hadamard designs is that their rows and columns are orthogonal.

The simplest Hadamard is $H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$. Larger Hadamard matrices may be generated as the Knonecker product of $\mathbf{H_2}$.with itself. For example, $\mathbf{H_4} = \mathbf{H_2} \otimes \mathbf{H_2}$, where $\otimes$ is the Knorneker product. Designs $\mathbf{H_8}$, $\mathbf{H_{16}}$, and so forth may be calculated recursively. Multiplying all components by –1 and appending the result to the bottom of the original design generate the foldover of the design.

In general, if the initial design is a main effects design, the design resulting from appending its foldover is an availability design. When the initial design is a Hadamard design, dropping the appropriate number of columns from the next largest available Hadamard provides designs for intermediate numbers of alternatives. Furthermore, columns of the initial Hadamard design may be multiplied by –1's to give a new Hadamard design that when augmented by its foldover will be an availability design with different set sizes for its choice sets.

Table 2

Constructing Choice and Estimation Designs

(A) Design H_4	**(B) Foldover D of H_4**
-1 -1 -1 1 1 -1 1 1 1 1 -1 1 -1 1 1 1 See [7] for discussion and illustration of how to create design matrices **H**.	-1 -1 -1 1 1 -1 1 1 1 1 -1 1 -1 1 1 1 1 1 1 -1 -1 1 -1 -1 -1 -1 1 -1 1 -1 -1 -1
(C) Effect Coding Equation 4	**(D) Dummy Coding for Design**
R1 ={-0.5 -0.5 -0.5 0.5} R2 = {0.5 -0.5 0.5 0.5} R3 = {0.5 0.5 -0.5 0.5} R4 ={-0.5 0.5 0.5 0.5} R5 = {0.5 0.5 0.5 -0.5} R6 = {0.5 0.5 -0.5 -0.5} R7 ={-0.5 -0.5 0.5 -0.5} R8 = {0.5 -0.5 -0.5 -0.5}	R1* = +0 0 0 1+ R2* = {1 0 1 1} R3* = {1 1 0 1} R4* = {0 1 1 1} R5* = {1 1 1 0} R6* = {1 1 0 0} R7* = {0 0 1 0} R8* = {1 0 0 0}
(E) Coding for Estimation Matrix (Effects) First Two Choice Sets	**(F) Coding for Estimation Matrix (Dummy) First Two Choice Sets**
0 0 0 R1 D R2 0 0 0 A 0 0 R2 0 C 0 0 0 R2 D R2 0 R2 0 AC R2 0 0 R2 AD 0 0 R2 R2 CD R2 0 R2 R2 ACD	0 0 0 R1* R2* 0 0 0 0 0 R2* 0 0 0 0 R2* R2* 0 R2* 0 R2* 0 0 R2* 0 0 R2* R2* R2* 0 R2* R2*

Table 2 illustrates how designs **D** for cross effects models may be generated. Panel (A) shows a foldover of the Hadamard constructed as **H4** = $\mathbf{H_2} \otimes \mathbf{H_2}$, and then multiplying the first column by –1. Panel (B) shows the complete availability design consisting of the foldover shown in panel (a) augmented by **H4**. The complete design consists of eight choice sets; four with a single alternative and four with three alternatives. The first row of Panel (B) provides the coding for Equation 4a and the second row the coding for Equation 4b. Panel (c) provides the coding that would be used if one estimates using the alternative coding discussed earlier in the paper, i.e., the elements of panel (a) are simply multiplied by 0.50. Panel (d) provides the traditional coding for a logit analysis using dummy variable coding. The –1's shown in panel (b) are simply replaced with zeros. Table 1 illustrates that the different coding schemes yield identical results.

Panels (e) and (f) show that the matrix used in estimating cross effects may be easily constructed from the appropriate design D. In the present case, there are four alternatives and the design matrix D has four columns indicating presence or absence of the alternatives. The estimation matrix will have four sets of four columns, for a total of 16. The first four columns represent the own effect and cross effects of other alternatives on the first alternative. The second set of four columns represent the own effect and cross effects of the alternatives on the second alternative, and so forth. For example, the first row of panel (e) has three blocks of four zeros indicating alternatives A, B, and C are absent from the design. The fourth position has row 1 (R1) of panel (c) inserted indicating the presence of alternative D in the set. Not that the coding of R1 has coding indicating alternative D is being chosen in the absence of alternatives A, B, and C.

The construction for the second choice set is slightly more complicated. It has three rows with row 2 (R2) of panel (c) inserted in the appropriate columns to indicate choice of the single options A, C, or D, with zeros elsewhere. It also has row R2 inserted twice in appropriate columns for coding of combinations of two alternatives being selected (the utility of the combination is the sum of the utilities of the single options.). The columns for the absent alternatives are represented with blocks of zeros. Finally, the combination of three alternatives being selected is represented with three blocks of R2 and one block of zeros representing the alternative missing; from the choice set. The remaining choice sets are coded following the same logic. The dependent variable in the subsequent analysis is the vector of logits of the probabilities of choosing the respective combinations in the choices set with the no choice in the set taken as the base.

A Hypothetical Example

Consider a university student who after a hard night of studying wants to get a pizza from a popular off-campus pizza restaurant. The restaurant does not deliver, but is only 20 minutes away walking, 10 minutes away by bike, 15 minutes away by bus (excluding waiting time), and 5 minutes away by automobile. A hypothetical representation of own and cross effects for this problem may be represented as Table 3.

Table 3

Values of Modes and Cross Effects of Other Modes.

	Walk	**Bike**	**Bus**	**Drive**
Walk	**-0.14**	**0.42**	**0.20**	**-2.14**
Bike	**-0.67**	**-0.77**	**-0.52**	**-0.96**
Bus	**-0.53**	**0.10**	**-5.34**	**-5.31**
Drive	**-0.03**	**0.10**	**-0.24**	**2.02**

The main diagonal of Table 3 has the "own effects" of the respective of the three modes available to the student. They reflect the average attractiveness of the mode across all combinations and choice sets in which it appears. The values in a given row of the table indicate the affect the other modes have on the row mode when both are in a choice set. That is, they may increase or decrease the average attractiveness of the mode depending on whether they are available in the choice set. For example, the value of 0.42 in the first row indicates that the option traveling by bicycle increases the attractiveness of the walking option. This may occur for example because the possibility of walking one way and bicycling the other is attractive to the student, perhaps to get some exercise before eating the pizza. The corresponding cross effect of bike on bicycle is negative (-0.67 in the second row). As discussed in connection with Figure 1, the net effect of the two coefficients is to increase the probability of the walking-bike combination and reduce the probability of the bike only choice.

Table 4

Mode Shares for Combinations Using Values in Table 3

Single Mode Choice Sets			
(a) P(Walk) 0.80	(b) P(Bike) 0.80	(c) P(Bus) 0.60	(d) P(Drive) 0.90
Choice Set Including Car (e)		Choice Set Excuding Car (f)	
Walk	0.01	Bike	0.42
Bus	0.00	Bus	0.01
Drive	0.77	Drive	0.02
Walk/Bus	0.00	Bike/Bus	0.15
Walk/Drive	0.06	Bike/Drive	0.26
Bus/Drive	0.00	Bus/Drive	0.01
Walk/Bus/Drive	0.00	Bike/Bus/Dive	0.10

Looking at the third row of Table 3, the cross effect of Driving on Bus of –5.31 substantially reduces the probability of choosing the Bus option when Driving is an option. The cross effect of Bus on Driving is –0.24, which modestly reduces the probability of driving. The net effect of the pattern of own and cross effects is shown in Table 4. Panels (a) –(d) give the mode shares when the alternative is to use a single mode or not get a pizza. Driving is the most attractive option, Walking and Biking are second, and using the Bus is the third most attractive option. Panel (e) gives the mode shares when the options are walking, busing, driving, or combinations of the three modes. Driving remains the dominant mode. Panel (f) gives mode shares in a choice set where Driving is not an option. Options involving biking or combinations of biking with other modes are the dominant choices in this choice set.

Response formats

In the hypothetical example, we assumed respondents could choose any combination that could be formed from the elements of a choice set. In fact, there are a wide number of response formats that could be used. The following formats can in general be applied:

1. unconstrained: respondents compose the portfolio from which choice sets will be selected.
2. pick 1 from each choice set
3. pick precisely n from each choice set of size t_i, $n \leq t_i$
4. pick precisely n of N ($n \leq N \leq t_i$)
5. pick up to a maximum of n ($n \leq N \leq t_i$)

Technically, the choice of response format does not have any bearing on the general modeling approach, provided the respondents can pick at least 2 alternatives. However, the amount of information differs. Allowing subjects to pick more than 1 tends to increase the number of observations as 2^n. One consequence may be infrequently chosen combinations, implying that ceteris paribus the reliability of predicting such portfolios will increase. However, it may also change the behavioral process, and as such the decision which response format to apply should mimic the actual choice process under investigation: in some situations individuals can only choose a single alternative (e.g., one transport mode at the same time), in some situations individuals may choose more options (e.g. choice of options in tourism travel), but not necessarily all. The “pick n from each choice set” and “pick n of N ($n \leq N \leq t_i$)” are of particular interest when one wishes to estimate supply elasticities.

Possible extensions

In the previous sections, we have discussed a basic approach to modeling portfolio choice. The simple example only concerned an individual’s choice of some combination of alternatives. To indicate the potential value of this approach, in this section we will discuss possible extensions.

Context

Many choice problems in transportation research are *conditional* on one or more contextual variables. Examples of such context variables are day of the week, weather conditions, time of day, spatial context, etc. Including context in portfolio choice implies that the suggested approach for constructing an experimental design to vary choice set composition needs to be nested under an orthogonal fractional factorial design, systematically varying the conditions. Context effects can then be represented as either main effects or interactions with the cross-effects [14]. A likely strategy will be to model the context effects as context cross effects on the alternative own effect and context by alternative cross effect interactions.

Sequences

Until now, we have assumed that no particular sequence is relevant for the choice problem. However, if may be that portfolio choice at one particular moment in time is *contingent* on the choice during the previous period of time. Examples in transportation would be activity program choice on weekend as a function of activity program choice on previous weekdays, or trip-chaining in an activity context. To study such problems, an appropriate task needs to be formulated. For example, subjects could be asked to pick a set of activities for the weekdays and then dependent on this choice, pick an activity agenda for the weekend. No specific requirements are need with respect to the design of the experiments and the estimation. However, the complexity of the model in terms of the number of effects will rapidly increase having an impact on the design of the study and experiment.

Group decision

Two perspectives on group decision making can be distinguished. In one case, the researcher is only interested in the outcomes of a group decision. The choice task can then be completed by the group (e.g., household) and everything else remains the same. This implies that the group is used as the unit of analysis. Perhaps the more interesting case however is that the group decision process has ramifications for individual choices. Individual have different preferences and perhaps different constraints that need to be negotiated. In this case, the experiments need to reflect this decision process. However, in principle, approaches to study group decision making for standard conjoint tasks [e.g.,15] can also be applied to the problem of group portfolio choice. For example, individual group members should first complete a portfolio decision task. Subsequently, the group at large should complete the task. Possible differences in main and cross effects could then be analyzed in terms of preference shifts as a result of the negotiation process. In addition to the cross effects, such models also allow one to estimate the relative influence of the group members on the ultimate portfolio choice.

Constraints

Various types of constraints may act on probabilities of portfolio choice. Some may be exogenous, while others may be endogenous. For example, portfolio choice may be subject to a time constraint. In this situation, the total should by definition sum to the available time budget. This problem is a redistribution problem. The effects should then add up to a constant. Portfolio choice may also be subject to a budget constraint. This situation is slightly different from the time allocation problem in the sense that portfolio choice may also involve a cost lower than the budget constraints, putting a different requirement on the effects to be estimated. An example of an endogenous constraint would be a tolerance threshold to travel. Exogenous constraints need to be reflected in the experimental task. For example, subjects can only choose those portfolios the costs of which are lower than or equal to the available budget. In the case of time allocation, the time involved in the portfolio choice should be equal to the available time budget.

Attributes

The most general case, which may include the above aspects as well, is the case where the choice of some portfolio, also depends on the attributes of the choice alternatives. Costs would be a typical example. In this case, different designs are required. There are two sets of cross effects with such designs, the availability effects discussed in this paper, and cross attribute own and cross effects. For example, the own attribute effect would be the average of a price level on the attractiveness of a product. Price attribute cross effects would be the effects of competitors prices on the own effect of a price level. The number of parameters to estimate increase rapidly with such designs. For example, with an availability design there are m^2 own and cross effects where m is the number of alternatives. With a two level price attribute, there could be $2*m^2$ parameters. Lazzari and Anderson [16] discuss one approach for estimating the cross effects of price with a single attribute. Wiley [17] discusses some generalizations of their approach to more levels.

Conclusions and discussion

The aim of this paper has been to draw the attention to the problem of portfolio choice and to suggest a general approach of how to model such choice problems in the context of conjoint analysis (stated choice methods). It has been shown how cross-effects may be used to estimate the effect of the composition of the chosen set of alternatives on the probability of choosing a particular portfolio, and that the problem can be represented in terms of random utility theory and conventional logit models. Design catalogs and design strategies have been developed for estimating cross effects.

We hope that this general discussion and the list of a wide variety of choice problems in transportation research that can be conceptualized and modeled as a portfolio choice will trigger empirical research on this class of decision problem. That would increase our experience with regard to various operational issues in the administration and implementation of these experiments. Issues that need pertinent consideration in this context are the symmetry or non-symmetry of the cross-effects, the consistent treatment of constraints (especially for group resource allocation problems with distribution effects, implying that constraints should be met at the individual level, the group level and the resource level), and inclusion of heterogeneity.

References

[1] Lam, S-H and Xie, F (2002), Transit Path-Choice Models That Use Revealed Preference And Stated Preference Data, *Transportation Research Record*, 1799, 58-65.

[2] Liu, X., Guan, H., Qin, H. (2006), Modeling Park-and-Ride Choice Behavior Based on Stated-Preference Survey, Proceedings TRB Annual Meeting, Washington, D.C.

[3] Whitaker, B., Terzis, G., Soong, E., and Yeh, W. (2005), Stated Preference as a Tool to Evaluate Airline Passenger Preferences and Priorities, *Transportation Research Record*, 1915, 55-61.

[4] Adler, Th., Falzarano, C.S., and Spitz, G. (2005), Modeling Service Trade-Offs in Air Itinerary Choices, *Transportation Research Record*, 1915, 55-61, 20-26.

[5] Louviere, J.J., Hensher, D.A. and J. Swait (2000), *Stated Choice Methods. Cambridge*, University Press, Cambridge, Mass.

[6] Hensher, D.A. (1994), Stated Preference Analysis of Travel Choices: The State of. Practice, *Transportation*, 21, pp. 107 -0.533.

[7] Anderson, D.A. and Wiley, J.B. (1992), Efficient Choice Set Designs for Estimating Availability Cross-Effects Models, *Marketing Letters*, 3, 357-70.

[8] Raghavarao, D. and Wiley, J.B. (1987), Testing Competing Effects Among Soft Drink Brands", C.E. McCulloch, S.J. Schwayer, G. Casella and S.R. Searle (ed.), *Statistical Design Theory and Practice*, Ithaca, N.Y., Cornell University Press, pp. 161-76.

[9] Raghavarao, D., Federer, W.T.and Schwager, S.J. (1986), Characteristics For Distinguishing Balanced Incomplete Block Designs With Repeated Blocks, *Journal of Statistical Planning and Inference*, 13, 151 -0.563.

[10] Calvin, L.D. (1954), Doubly Balanced Incomplete Block Designs for Experiments in Which The Treatment Effects Are Correlated, *Biometrics*, 10, 61-88.

[11] Bhaumik, D.K. (1995), Optimality in the Competing Effects Model, *Sankhya-B*, 57, 48-56.

[12] Anderson, D., Borgers, A.W.J., Ettema, D. and Timmermans, H.J.P. (1992), Estimating Availability Effects in Travel Choice Modeling: A Stated Choice Approach. *Transportation Research Record*, no. **1357**, pp. 51-65.

[13] Raghavarao, D. and Zhou, B. (1998), Universal Optimality of UE 3-Designs for a Competing Effects Model," Communications in Statistics: Theory Methods, 27, 1 pp. 153 -0.564.

[14] Oppewal, H. and Timmermans, H.J.P. (1991), Context Effects and Decompositional Choice Modelling. *Papers in Regional Science*, **70**, pp 113-153.

[15] Molin, E.J.E., Oppewal, H. and Timmermans, H.J.P. (2000), A Comparison of Full Profile and Hierarchical Information Conjoint Methods in Modeling Group Preferences, *Marketing Letters*, **11**, 165 -172.

[16] Lazari, A.G. and D. A. Anderson (1994), Designs of Discrete Choice Set Experiments of Estimating Both Attribute and Availability Cross Effects, *Journal of Marketing Research*, Vol. 31, No. 3 (Aug., 1994) , pp. 375-383

[17] Wiley, J.B. (2001), Experimental Designs in Choice Experiments, in *Recent Advances in Design of Experiments and Related Areas*, Nova Science Publishers