CHARLES F. MANSKI*

THE STRUCTURE OF RANDOM UTILITY MODELS

INTRODUCTION

This paper presents a formal analysis of the distributional structure of random utility models. Its essential conceptual contribution lies in its explicit treatment of the processes making perfect predictions of choice behavior unattainable. The paper's technical contribution is its demonstration of the manner in which these processes induce the distributional properties of a model. In contrast to the approach adopted here, the existing literature on random utility models generally imposes distributional assumptions directly. Because it leaves so much implicit, this practice has often caused researchers to remain unaware of the restrictiveness of their models. Only a few authors, working with specific simple models, have chosen to lay the foundations of their research more deeply. Their efforts will be described shortly.

Historically, random utility models, a subset of the class of probabilistic choice models, were first developed by psychologists in the attempt to characterize observed inconsistencies in patterns of individual behavior.[1] A definition offered by Block and Marschak (1960) can be paraphrased for our purposes as follows:

> Let $\mathfrak{a}$ be a finite set of alternatives, T be a finite population of decision makers and let $\in^c$ mean 'is chosen from'. Then choice is consistent with a random utility model if there exist real valued random variables U_{at}, all $a \in \mathfrak{a}$, $t \in T$ such that $\Pr_t(a \in^c C) = \Pr(U_{at} \geqslant U_{a't}$, all $a' \in C)$ for all alternatives $a \in C$, all non-null choice sets $C \subset \mathfrak{a}$ and all decision makers $t \in T$.

Later, economists, beginning with McFadden (1968) embraced such models as an econometric representation of maximizing behavior. In this formulation, utilities are treated as random variables not to reflect a lack of rationality in the decision maker but to reflect a lack of information regarding the characteristics of alternatives and/or decision makers on the part of the observer.[2]

Theory and Decision **8** (1977) 229–254.

Differing substantive interpretations not withstanding, both the econometric and psychometric literatures focussed early on one simple specification for the random utility function, the form $U_{at} = V_{at} + \epsilon_{at}$ where $(V_{at}, a \in \mathfrak{a}, t \in T)$ are a set of constants and $(\epsilon_{at}, a \in \mathfrak{a}, t \in T)$ are a set of independent and identically distributed random variables. Among such 'independent and identically distributed random utilities' (IIDRU) models, the Luce (1959, 1965) and McFadden (1968, 1973) model undoubtedly has the best analytical and computational properties. Domenchic and McFadden (1975) offer other examples of IIDRU models.

Although empirical practice still confines itself to the exclusive use of IIDRU models, the special nature of such models has long been recognized. Beginning in Debreu's (1960) review of Luce's pioneering work (Luce, 1959), examples of situations in which IIDRU models yield counter-intuitive behavioral forecasts have become numerous.[3] Concern with the development of tractable but less restrictive models has, to a degree, increased as well. Two such models, Tversky's elimination-by-aspects model (Tversky, 1972a, 1972b) and a model constructed by Quandt and Young (1969) and generalized by Domenchic and McFadden (1975) are now available. Both models derive their distributional properties from more basic assumptions, both behave more intuitively than do IIDRU models yet both contain an IIDRU model as a special case.

While the above developments have contributed in important ways to our grasp of the structure of random utility models, our understanding of such models remains fragmentary and in many ways superficial. Our continuing inability to delineate the appropriate domain for IIDRU models typefies the situation. Instead of turning to formal analysis of the problem, discussion has become mired in the search for ad hoc corrections to be applied to existing models so as to make them behave more plausibly. An ill-defined dichotomy between 'similar' alternatives, for which IIDRU models are supposed to be inappropriate and 'distinct' alternatives, which such models are said to suitably handle, has also arisen.[4]

Repeated frustration in my own attempts to operationalize the similar-distinct dichotomy gradually convinced me that further progress requires not new technique but rather a reformulation of the random utility model at a position logically prior to that hitherto adopted. That reformulation, in terms of the processes making perfect choice predictions unattainable, is contained in Part I of the present paper. Part II then examines the distributional properties of random utility models.

In order not to raise unjustified hopes to applied researchers, I must point out that this paper's scope is limited to theoretical issues. Questions of operationalizing the models presented in terms of tractable estimation and forecasting procedures are, by and large, not considered.[5]

I. THE GENERAL MODEL

This development of the random utility model begins with a statement of primitive concepts and definitions. Next, the decision rule governing the behavior of a given decision maker faced with a specified choice set is specified. Then the process associating decision makers with choice sets is introduced.

At this point, attention shifts to the observer. I detail the limited information regarding the choice problem generating process, the decision rule governing choice, and the characteristics of decision makers and alternatives – he possesses. Finally, I show how lack of complete information induces the random utility model and its associated choice probabilities.

Mathematically, this paper uses only the concepts of elementary set and probability theory. Nevertheless, the proper interpretation of certain assumptions and results can be quite subtle. Throughout the model's development, I attempt to point our where steps are restrictive and to interpret their import.[6]

A. *Primitive Concepts and Definitions*

Let *choice* be an operation mapping a set into one of its non-null elements and a *decision maker* be an agent performing choice operations according to a fixed rule. A finite population of decision makers, designated T, is assumed to exist. Assume the existence of a finite set $\mathfrak{a}$, termed the *alternative space*, such that the choice operation is defined over each of $\mathfrak{a}$'s non-null subsets. The elements of $\mathfrak{a}$ are then termed *alternatives*, the ordered non-null subsets of $\mathfrak{a}$ are termed *choice sets* and the set of all ordered non-null subsets of $\mathfrak{a}$ is termed the *choice set space*, designated Γ.

DISCUSSION: In the above, choice sets are defined as ordered sets for analytical reasons. To the utility maximizing decision maker, the ordering of the elements of such a set is irrelevant.

B. *The Decision Rule*

Let U_t be a real valued function defined over $\mathfrak{a}$ and termed a ***utility function***. Faced with any choice set $C \in \Gamma$, decision maker t is assumed to select an $a \in C$ such that $U_{at} \geqslant U_{a't}$, all $a' \in C$.[7]

Let X and S be finite subsets of real spaces and $\mathfrak{a} \to X$, $T \to S$ be mappings associating every $a \in \mathscr{A}$ with an $x_a \in X$, every $t \in T$ with an $s_t \in S$. X and S are ***attribute*** representations of the alternative space $\mathfrak{a}$ and decision maker space T while x_a and s_t are attribute vectors for a and t respectively. The attribute representations X and S are said to be ***utility relevant*** if there exists a real valued function w, defined over $X \times S$ such that

$$U_{at} = w(x_a, s_t) \quad \text{all} \quad a \in \mathfrak{a}, \quad t \in T.$$

DISCUSSION: (1) If X is an attribute representation for $\mathfrak{a}$, then the choice set space Γ has a corresponding representation, namely the class of ordered sets $R = (r_c, C \in \Gamma)$ where $r_c = [(x_a, \text{all } a \in C)]$. If the mapping $\mathfrak{a} \to X$ is one-to-one, then R is the set of all non-null subsets of X.[8]

(2) We shall, in this paper, be interested only in utility relevant representations for $\mathfrak{a}$ and T.

In general, each $\mathfrak{a}$ and T admit infinitely many such representations and the observer must select among them. While often this choice is merely one of normalization, sometimes it is less trivial. For example, the utility of a transportation mode might be specified to be a function of its physical characteristics or of its 'qualities', comfort, convenience, privacy etc., appropriately scaled. Both types of specification are found in the literature. The observer's selection of a representation will depend on the nature of his attribute data and on his ability to specify the w function associated with alternative representations.

Our concern will be with the derivation of a random utility model from a given representation, not with the choice among representations.

(3) In establishing real valued attribute representations for the primitive sets $\mathfrak{a}$ and T, it is natural to inquire what properties of real spaces are being exploited. At this stage, these representations are merely formalisms. Even later, the only property used is the expressibility of a real space as a Cartesian product set.[9]

Nowhere are X and S viewed as generating linear spaces. In fact, linear combinations of elements in X need correspond not only to no element of $\mathfrak{a}$ but to no logically possible alternative. For example, in mode choice models travel time must be positive and the number of occupants in a vehicle is confined to the positive integers.[10]

C. *The Generation of Choice Problems*

The framework of classical choice theory, within which the present effort lies, explains individual behavior as the outcome of a two step recursive process. First, exogeneous forces pose a choice problem, that is an individual and an associated choice set. Then, with the choice set well-defined, the individual selects among the available alternatives.

In general, research on choice theory has focused on the second stage process, the characterization of classes of decision rules, formalization of choice set structure and study of the properties of the outcome when decision rules of a given class are applied to choice sets of a specified structure.[11] In the present development of the random utility model, the description of the first stage, the mechanism generating choice problems, occupies an equally central position. Having already assumed a decision rule, utility maximization, and a choice set structure, finite point sets, we now describe the generation of choice problems, first formally and then through an interpretive discussion.

Let a *choice problem* be a pair (C, t) drawn from the product space $\Gamma \times T$ according to a probability measure $M_{\Gamma T}$, defined over $\Gamma \times T$. $M_{\Gamma T}$ is the *choice problem generating process*.

From $M_{\Gamma T}$, a variety of useful probability measures may be derived. These include the following:

M_T, a distribution with domain T and defined by $M_T(t) = \Sigma_{C \in \Gamma} M_{\Gamma T}(C, t)$ is the *decision maker generating process.*

$M_\Gamma(C/t)$, $t \in T$, a set of distributions each with domain Γ and defined by $M_\Gamma(C/t) = M_{\Gamma T}(C, t)/M_T(t)$ are the *choice set generating processes.*

Because the ordering of elements within a choice set is arbitrary, we shall adopt the convention that $M_{\Gamma T}(C, t) = M_{\Gamma T}(\tilde{C}, t)$ whenever C and $\tilde{C}$ are identical except for order.

DISCUSSION: Specification of a distribution $M_{\Gamma T}$ and a set of utility values U_{at}, $a \in \mathfrak{a}$, $t \in T$ fully characterizes a classical choice process. The conciseness of this formalization is both pleasing and powerful. At the same time, however, it is implicitly restrictive and undeniably unintuitive. The discussion below attempts to provide the intuition for and make explicit the restrictions in this model of choice problem generation.

(1) Noting that $M_{\Gamma T}(C, t) = M_{\Gamma}(C \mid t) M_T(t)$, consider first the generation of choice sets, as embodied in $M_{\Gamma}(C \mid t)$. This process is most easily interpreted when there exist a set of decision makers TT who control the alternatives $\mathfrak{a}$. The generation of choice sets for T is then the outcome of a set of prior decisions made by the members of TT. These are the (not necessarily independent) binary decisions specifying whether each $a \in \mathfrak{a}$ will or will not be made available to each $t \in T$.

For example, if T is a set of students and $\mathfrak{a}$ is a set of colleges, then college choice sets are generated for T through the admissions decisions of the intitutional administrators TT controlling $\mathfrak{a}$. Similarly, T might be a set of workers, $\mathfrak{a}$ a set of jobs and TT a set of firms making job offers. Or T might be the set of draft eligible males, $\mathfrak{a}$, a set of military and civilian alternatives and TT 'nature' operating through the draft lottery to restrict the availability of civilian alternatives.

Turn now to the process generating decision makers, M_T. In the context of random utility function estimation, where the observer draws a sample from T and then notes their choices, interpretation of M_T as describing the sampling process is straightforward. A different interpretation of M_T derives from the psychological model of random utilities. There T specifies a set of decision rules from which the individual draws, according to the distribution M_T, when faced with a choice situation.[12]

(2) The classical assumption of a recursive choice process often restricts its empirical utility. In particular, when TT is a set of conscious decision makers (as opposed to 'nature'), symmetric treatment of T and TT seems more appropriate. A recursive structure gains attractiveness if it allows recontracting. While the present development of the random utility model is static in the sense that it describes only one move by each actor, a sequence of such models could conceivably be linked so as to describe a recontracting process. Such a sequence might capture the essence of game theoretic models while retaining the analytically useful concept of a choice set.[13]

(3) While we have presented $M_{\Gamma T}$ as a distribution characterizing a single choice problem drawn from $\Gamma \times T$, greater interest lies in describing a sequence of such draws. For this purpose, $M_{\Gamma T}$ continues to suffice only if choice problems are generated independently. More generally, if $(C, t)_n$, $n = 1, \ldots N$ were a sequence of choice problems, a probability measure over $(\Gamma \times T)^N$ would be required.[14]

D. *The Observer's Information Base*

We now postulate an *observer*, external to the choice process described by $M_{\Gamma T}$ and U, and endowed with limited information regarding that process. The present task, the final step in developing the random utility model, is to specify the observer's information base.

We assume the following:

(1) The observer has no direct knowledge of the utility values U_{at}, $a \in \mathfrak{a}$, $t \in T$.

(2) There exist utility relevant attribute representations X and S with an associated utility function w such that for every specified $x \in X$, $\mathrm{a} \in S$, $w(x, \mathrm{s})$ is known.

(3) The attribute values x_a, $a \in \mathfrak{a}$ and s_t, $t \in T$ are only partially known to the observer. In particular, there exists a partitioning of X, defined by $X = [X_o{:}X_u]$ and one of S, $S = [S_o{:}S_u]$ such that the subvectors $X_o = (x_{ao}, a \in \mathfrak{a})$ and $S_o = (s_{to}, t \in T)$ are observed without error but $X_u = (x_{au}, a \in \mathfrak{a})$ and $S_u = (s_{tu}, \mathrm{t} \in T)$ are unobserved.

(4) For every choice problem (C, t), the probability, conditioned on knowledge of the observed attributes $r_{co} = (x_{ao}, a \in C)$ and s_{to}, that the unobserved attributes $r_{cu} = (x_{au}, a \in C)$ and s_{tu} take on any values $\bar{r}_{cu}$ and $\bar{s}_{tu}$ is known. This probability is given by

$$\Pr(\bar{r}_{cu}, \bar{s}_{tu} \mid r_{co}, s_{to}) = \Pr(\bar{r}_c, \bar{s}_t) \,/\, \Pr(r_{co}, s_{to}) =$$

$$= \sum_{\substack{(\tilde{C}, \tilde{t}):\, r_{\tilde{c}} = \bar{r}_c \\ s_{\tilde{t}} = \bar{s}_t}} M_{\Gamma T}(\tilde{C}, \tilde{t}) \Bigg/ \sum_{\substack{(\tilde{C}, \tilde{t}):\, r_{\tilde{c}o} = r_{co} \\ s_{\tilde{t}o} = s_{to}}} M_{\Gamma T}(\tilde{C}, \tilde{t}).^{15}$$

DISCUSSION: (1) Our informational assumptions place the locus of observer ignorance totally in his inability to observe certain attributes of alternatives and of decision makers. The resulting random utility model is consistent with the classical notion of probability. An informationally weaker model would permit incomplete knowledge of the structure of w and/or of the probabilities $\Pr(\bar{r}_{cu}, \bar{s}_{tu} \mid r_{co}, s_{to})$.[16] In such a model, these probabilities would be subjective and w a subjectively random function. Explicit Bayesian analysis of random utility models is beyond the scope of the present paper.[17]

(2) It will become apparent in the next section that the present develop-

ment of the random utility model does not formally require that the choice problem generating process $M_{\Gamma T}$ be known, only the derived probabilities $\Pr(\bar{r}_{cu}, \bar{s}_{tu} \mid r_{co}, s_{to})$. Unfortunately, the latter probabilities are very difficult to substantively interpret. Hence, we choose to work with the former ones. Calculation of 'outcome probabilities' in Section F does require knowledge of $M_{\Gamma T}$.

(3) In writing $X = [X_o : X_u]$ and $S = [S_o : S_u]$ we assume that X and S partition 'conformably' into observed and unobserved parts. That is, there exist no attributes which are observed for some alternatives (decision makers) but unobserved for others. This assumption, which is in no way crucial, simplifies the presentation.

(4) In the random utility literature, the function w is often written in the additive form $w(x, s) \equiv V(x_o, s_o) + \epsilon(x, s)$, where V captures that part of w additively separable in observed attributes and ϵ the residual. Note that V is a well-defined function only if X and S each partitions conformably into observed and unobserved parts.

E. *The Random Utility Model*

We now show that the decision rule, the choice problem generating process and the observational information assumed in Sections B, C, and D induce a random utility model describing behavior.

In every choice problem (C, t), the vector of utilities $W_{ct} = w((x_{ao}, x_{au}), (s_{to}, s_{tu}))$, $a \in C$ are functions of the unobserved attributes x_{au}, $a \in C$ and s_{tu}. Hence, the utilities are *observationally random* with the following distribution:

$$\Pr(\bar{W}_{ct} \mid r_{co}, s_{to}) = \Sigma_{(\bar{r}_c, \bar{s}_t): W = \bar{W}_{ct}} \Pr(\bar{r}_c, \bar{s}_t) / \Pr(r_{co}, s_{to})$$

$$= \frac{\Sigma_{\substack{(\tilde{C}, \tilde{t}): r_{\tilde{c}o} = r_{co} \\ s_{\tilde{t}o} = s_{to} \\ W_{\tilde{c}\tilde{t}} = \bar{W}_{ct}}} M_{\Gamma T}(\tilde{C}, \tilde{t})}{\Sigma_{\substack{(\tilde{C}, \tilde{t}): r_{\tilde{c}o} = r_{co} \\ s_{\tilde{t}o} = s_{to}}} M_{\Gamma T}(\tilde{C}, \tilde{t})}$$

Recall that for each choice problem $(C, t) \in \Gamma \times T$ and any $a \in C$, t chooses a from C if and only if $w((x_{ao}, x_{au}), (s_{to}, s_{tu})) \geqslant w((x_{\tilde{a}o}, x_{\tilde{a}u}), (s_{to}, s_{tu}))$, all

$\tilde{a} \in C$. Letting $w_{at} = w(x_a, s_t)$, it follows that observationally

$$\Pr_t(a \in^c C) = \Sigma_{\bar{W}_{ct}} : \bar{w}_{at} \geqslant \bar{w}_{\tilde{a}t}, \tilde{a} \in C \Pr(\bar{W}_{ct} \mid r_{co}, s_{to}).$$

That is, choice is consistent with the random utility model.[18]

DISCUSSION: An example may help to summarize the model. Let

$$\mathfrak{a} = (\alpha, \beta, \gamma) \qquad\qquad T = (\sigma, \tau)$$

$$x = \begin{array}{c|ccc} & \alpha & \beta & \gamma \\ \hline x_1 & 1 & 1 & 2 \\ x_2 & \frac{1}{2} & -\frac{1}{2} & -1 \end{array} \qquad\qquad S = \begin{array}{c|cc} & \sigma & \tau \\ \hline s & 1 & 0 \end{array}$$

That is, there are three alternatives, α, β, and γ each characterized by the two real valued attributes x_1 and x_2 and there are two decision makers σ and τ each characterized by the one real attribute s.

Assume that the utility function governing behavior is $w(x, s) = x_1 + x_2\, s$.

Furthermore, let the choice problem generating process be defined by the following probabilities:

$M_{\Gamma T}\ [(\alpha, \beta, \gamma), \sigma] = \frac{2}{36}$	for each of the six ordered choice sets whose elements are α, β, and γ.
$M_{\Gamma T}\ [(\alpha, \beta), \sigma] = \frac{1}{12}$	for each of the two ordered choice sets whose elements are α and β.
$M_{\Gamma T}\ [(\alpha, \beta, \gamma), \tau] = \frac{1}{36}$	for each of the six ordered choice sets whose elements are α, β and γ.
$M_{\Gamma T}\ [(\alpha, \beta), \tau] = \frac{2}{12}$	for each of the two ordered choice sets whose elements are α and β.

If x_1 and s are observed but x_2 is not, then it can be deduced that for any ordered choice set composed of α, β and γ,

$$\Pr_\sigma(\alpha \in^c (\alpha, \beta, \gamma)) = \tfrac{1}{2} \qquad \Pr_\tau(\alpha \in^c (\alpha, \beta, \gamma)) = 0$$

$$\Pr_\sigma(\beta \in^c (\alpha, \beta, \gamma)) = \tfrac{1}{2} \qquad \Pr_\tau(\beta \in^c (\alpha, \beta, \gamma)) = 0$$

$$\Pr_\sigma(\gamma \in^c (\alpha, \beta, \gamma)) = 0 \qquad \Pr_\tau(\gamma \in^c (\alpha, \beta, \gamma)) = 1.$$

However, if s as well as x_2 are unobserved, then

$$\Pr_\sigma(\alpha \in^c (\alpha, \beta, \gamma)) = \Pr_\tau(\alpha \in^c (\alpha, \beta, \gamma)) = \tfrac{1}{3}$$
$$\Pr_\sigma(\beta \in^c (\alpha, \beta, \gamma)) = \Pr_\tau(\beta \in^c (\alpha, \beta, \gamma)) = \tfrac{1}{3}$$
$$\Pr_\sigma(\gamma \in^c (\alpha, \beta, \gamma)) = \Pr_\tau(\gamma \in^c (\alpha, \beta, \gamma)) = \tfrac{1}{3}.$$

F. *The Outcome of the Choice Process*

The descriptive power of a random utility model is contained in the set of *choice probabilities* $[\Pr_t(a \in^c C), \text{all } a \in C \in \Gamma, \text{all } t \in T]$ it induces. Although the preceding derivation of these probabilities required specification of both steps in the classical choice process, the probabilities themselves describe only the outcome of the second step. That is, they characterize choice from a well-defined choice set by a given decision maker.

Often, ultimate interest rests in obtaining a description of the outcome of the entire choice process. The following *outcome probabilities* provide this description:

Let $\Pr(a)$ designate the probability that a is the outcome of the choice process. Then, for every $a \in \mathfrak{a}$,

$$\Pr(a) = \Sigma_{C \in \Gamma} \Sigma_{t \in T} \Pr_t(a \in^c C) M_{\Gamma T}(C, t).$$

DISCUSSION: (1) Where the concern is to forecast aggregate market shares, for example transportation modal splits and relative sales volumes of competing manufacturers, the outcome probabilities are sufficient statistics. Sometimes, it is desired to forecast the composition as well as the aggregate levels of outcomes. For example, we may wish to forecast the racial composition of college enrollments or the socio-economic patterns in residential location. To provide these forecasts, we need simply partition $\mathfrak{a}$ and T into the categories of interest and construct the relevant probabilities. Formally,

Let

$$\mathscr{A} = \bigcup_{n=1}^{N} \mathfrak{a}_n, T = \bigcup_{m=1}^{M} T_m,$$

where

$$\mathfrak{a}_n \cap \mathfrak{a}_{n'} = \phi \text{ if } n \neq n' \text{ and } T_m \cap T_{m'} = \phi \text{ if } m \neq m'.$$

Then

$$\Pr(a \in \mathfrak{a}_n, t \in T_m) = \Sigma_{a \in \mathfrak{a}_n} \Sigma_{t \in T_m} \Sigma_{C \in \Gamma} \Pr_t (a \in^c C) M_{\Gamma T}(C, t).$$

(2) Current methods for estimating the parameters of random utility functions require ex post observation of a sequence of choice problems for each of which the decision maker, choice set and chosen alternative are known. Often, however, the survey instrument used in estimation supplies the identities of the decision makers and his chosen alternative but not those of his feasible inferior alternatives. In such situations, the probability of the observation is

$$\Pr_t(\mathrm{a}) = \Sigma_{C \in \Gamma} \Pr_t(a \in^c C) M_\Gamma \,(C \mid t, a \in C),$$

where $M_\Gamma\,(C \mid t, a \in C)$ is the probability that C is drawn given that a must be an element of the realized choice set.

(3) Empirical application of outcome probabilities often presents a serious computational problem. That is, the summation $\Sigma_{C \in \Gamma}$ contains $2^{|\Gamma|} - 1$ terms, where $|\Gamma|$ is the order of Γ. As a result, it has become common practice in estimation studies to 'impute' the identities of unobserved inferior alternatives and view choices as if they had been made from the imputed choice sets. That is, if $\tilde{C}$ is the imputed set, the probability of the observation is taken to be $\Pr_t(a \in^c \tilde{C})$ rather than $\Pr_t(a)$ as defined above.

In the literature, two different imputation procedures have been used. Ben-Akiva and Lerman (1974), and other analysts of modal choice impute an 'expected choice set'.[19] Kohn *et al.* (1974) draw a choice set randomly from Γ according to the distribution $M_\Gamma\,(C \mid t, a \in C)$. Their choice probability $\Pr_t(a \in^c \tilde{C})$ is then an unbiased estimate of $\Pr_t(a)$. As yet, no general analysis of the consequences of alternative imputation rules has been attempted. McFadden and Reid (1974) and Westin (1974) do, however, show in a binary choice context that expected choice set imputation generates biased parameter estimates and biased forecasts.

II. SPECIAL MODELS

The work of Part I has demonstrated that if decision makers are utility maximizers and if the process associating decision makers with choice sets can be specified, then, to an observer possessing incomplete knowledge of the characteristics of decision makers and alternatives, a random utility model describes behavior. We now probe beneath this basic result to explain the formation of random utility models with special distributional properties.

Assume that the behavioral and informational conditions described earlier are met so that the statement

$$\Pr(\bar{W}_{ct} \mid r_{co}, s_{to}) = \sum_{\substack{(\tilde{C}, \tilde{t}): r_{\tilde{c}o} = r_{co} \\ s_{\tilde{t}o} = s_{to} \\ W_{\tilde{c}\tilde{t}} = W_{ct}}} M_{\Gamma T}(\tilde{C}, \tilde{t}) \Bigg/ \sum_{\substack{(\tilde{C}, \tilde{t}): r_{\tilde{c}o} = r_{co} \\ s_{\tilde{t}o} = s_{to}}} M_{\Gamma T}(\tilde{C}, \tilde{t})$$

characterizes observed choice behavior. Within the context of this general random utility model, the probability distribution of the random vector W_{ct} is determined by three factors, the form of the utility function $w(x, s)$, the nature of the choice problem generating process $M_{\Gamma T}$ and the extent of the observer's information which partitions the attributes $(x_a, a \in \mathfrak{a})$ and $(s_t, t \in T)$ into observed and unobserved parts.

We should like to demonstrate exactly how conditions placed on $w(x, s)$, $M_{\Gamma T}$, x_{ao} and s_{to} induce the distributional structure of random utility models. Equally strong is a desire to shed light on the substantive restrictions implied by the various particular models developed in the literature. To meet these dual objectives, the following analysis will focus on three familiar classes of random utility models, the independent random utilities, the independent and identically distributed random utilities and the random coefficients models. In what follows, each class of models is defined, theorems giving structural conditions inducing the class are presented and each theorem is discussed.

A. *The Independent Random Utility Models*

Consider a choice problem (C, t) drawn from $\Gamma \times T$. The associated random utility vector W_{ct} is said to have the *independent random utilities* (IRU) property if

$$\Pr(\bar{W}_{ct} \mid r_{co}, s_{to}) = \prod_{a \in C} \Pr(\bar{w}_{at} \mid x_{ao}, s_{to})$$

for all values of $\bar{W}_{ct}$.[20]

The theorem below gives conditions sufficient to ensure that a utility vector W_{ct} has the IRU property.

THEOREM: *Assume the following*:

(i) *for every*

$$(C, t) \in \Gamma \times T, M_\Gamma(C \mid t) = (\Pi_{a \in C} P(a \mid t))\,(\Pi_{\beta \notin C}\ (1-P(\beta \mid t)))$$

where

$$P(a \mid t) = \Sigma_{\tilde{C}:\ a \in \tilde{C}} M_{\Gamma T}(\tilde{C}, t) \Big/ \Sigma_{\tilde{C} \in \Gamma} M_{\Gamma T}(\tilde{C}, t)$$

is the probability that alternative a is an element of the choice set faced by decision maker t. Furthermore, for every t, $P(a \mid t) = 1$ for at least one $a \in \mathfrak{a}$.[21]

(ii) For every $(a, t) \in \mathfrak{a} \times T$, $P(a \mid t) = P(a \mid s_t)$. That is, decision makers whose utility relevant attributes are equal face equal probabilities $P(a \mid t)$.

(iii) *For every $t \in T$, $s_{to} = s_t$. That is, all utility relevant decision maker attributes are observed.*

Let (C, t) be a choice problem such that the observed attributes $(x_{ao}, a \in C)$ are distinct. Then, for all values of

$$\bar{W}_{ct}, \Pr(\bar{W}_{ct} \mid r_{co}, s_{to}) = \prod_{a \in C} \Pr(\bar{w}_{at} \mid x_{ao}, s_{to}).$$

Proof: Under the assumptions of the theorem,

$$\Sigma_{\substack{(\tilde{C}, \tilde{t}):\ r_{\tilde{c}o} = r_{co} \\ s_{\tilde{t}o} = s_{to} \\ W_{\tilde{c}\tilde{t}} = \bar{W}_{ct}}} M_{\Gamma T}(\tilde{C}, \tilde{t})$$

$$= \Sigma_{\substack{\tilde{C}:\ r_{\tilde{c}o} = r_{co} \\ W_{\tilde{c}t} = \bar{W}_{ct}}} \Sigma_{\tilde{t}:\ s_{\tilde{t}o} = s_{to}} M_\Gamma(\tilde{C} \mid \tilde{t})\, M_T(\tilde{t})$$

$$= \left[\Sigma_{\substack{\tilde{C}:\ r_{\tilde{c}o} = r_{co} \\ W_{\tilde{c}t} = \bar{W}_{ct}}} \left(\Pi_{\alpha \in \tilde{C}} P(\alpha \mid s_{to}) \right) \left(\Pi_{\substack{\beta:\beta \notin \tilde{C} \\ x_{\beta o} \in r_{co}}} \ (1-P(\beta \mid s_{to})) \right) \right.$$

$$\left. \times \left(\Pi_{\substack{\gamma:\gamma \notin \tilde{C} \\ x_{\gamma o} \notin r_{co}}} \ (1-P(\gamma \mid s_{to})) \right) \right] \cdot \left(\Sigma_{\tilde{t}:\ s_{\tilde{t}o} = s_{to}} M_T(\tilde{t}) \right)$$

$$= \left[\Sigma_{\tilde{C}:\, \substack{r_{\tilde{c}o} = r_{co} \\ W_{\tilde{c}t} = \bar{W}_{ct}}} \left(\Pi_{\alpha\in\tilde{C}} P(\alpha \mid s_{to})\right)\left(\Pi_{\substack{\beta:\beta\notin\tilde{C} \\ x_{\beta o}\in r_{co}}} (1-P(\beta \mid s_{to}))\right)\right]$$

$$\times \left(\Pi_{\gamma:x_{\gamma o}\notin r_{co}} (1-P(\gamma \mid s_{to}))\right) \left(\Sigma_{\tilde{t}:\, s_{\tilde{t}o}=s_{to}} M_T(\tilde{t})\right)$$

In the above, assumption (iii) guarantees the first equality while assumptions (i) and (ii) guarantee the second. The third equality rearranges the resulting expression.

By analogous transformations, it can be shown that

$$\Sigma_{(\tilde{C},\tilde{t})\, \substack{r_{\tilde{c}o} = r_{co} \\ s_{\tilde{t}o} = s_{to}}} M_{\Gamma T}(\tilde{C}, \tilde{t}) = \left[\Sigma_{\tilde{C}:r_{\tilde{c}o}=r_{co}} \left(\Pi_{\alpha\in\tilde{C}} P(\alpha \mid s_{to})\right)\right.$$

$$\left.\left(\Pi_{\substack{\beta:\,\beta\notin\tilde{C} \\ x_{\beta o}\in r_{co}}} (1-P(\beta \mid s_{to}))\right)\right]\left(\Pi_{\gamma:x_{\gamma o}\notin r_{co}} (1-P(\gamma \mid s_{to}))\right)$$

$$\times \left(\Sigma_{\tilde{t}:s_{\tilde{t}o}=s_{to}} M_T(\tilde{t}).\right)$$

Hence

$$\Pr(\bar{W}_{ct} \mid r_{co}, s_{to}) =$$

$$\frac{\Sigma_{\tilde{C}:\, \substack{r_{\tilde{c}o} = r_{co} \\ W_{\tilde{c}t} = \bar{W}_{ct}}} \left(\Pi_{\alpha\in\tilde{C}} P(\alpha \mid s_{to})\right)\left(\Pi_{\substack{\beta:\beta\notin\tilde{C} \\ x_{\beta o}\in r_{co}}} (1-P(\beta \mid s_{to}))\right)}{\Sigma_{\tilde{C}:r_{\tilde{c}o}=r_{co}} \left(\Pi_{\alpha\in\tilde{C}} P(\alpha \mid s_{to})\right)\left(\Pi_{\substack{\beta\notin\tilde{C} \\ x_{\beta o}\in r_{co}}} (1-P(\beta \mid s_{to}))\right)}$$

Observe that the class of choice sets

$$\left(\tilde{C}: \substack{r_{\tilde{c}o} = r_{co} \\ W_{\tilde{c}t} = \bar{W}_{ct}}\right) = \bigtimes_{a\in C} \left(\alpha: \substack{x_{\alpha o} = x_{ao} \\ w_{\alpha t} = w_{at}}\right)$$

and that the class

$$(\tilde{C}: r_{\tilde{c}o} = r_{co}) = \bigtimes_{a\in C} (\alpha: x_{\alpha o} = x_{ao})$$

where $\times_{a\in C}$ denotes the Cartesian product operator.

Additionally, observe that if $(x_{ao},\ a\in C)$ are distinct attribute vectors, then for any $\tilde{C}$: $r_{\tilde{c}o} = r_{co}$,

$$\left(\Pi_{\alpha\in\tilde{C}}\, P(\alpha \mid s_{to})\right)\left(\Pi_{\substack{\beta:\beta\notin\tilde{C}\\ x_{\beta o}\in r_{co}}} (1-P(\beta \mid s_{to}))\right)$$

$$= \Pi_{a\in C}\left(P(\alpha_a \mid s_{to})\,\Pi_{\substack{\beta:\beta\neq\alpha_a\\ x_{\beta o}=x_{ao}}} (1-P(\beta \mid s_{to}))\right),$$

where α_a is the unique element of $\tilde{C}$ such that $x_{\alpha_a o} = x_{ao}$. It follows from the above that

$$\Pr(W_{ct} \mid r_{co}, s_{to})$$

$$= \Pi_{a\in C}\left(\frac{\Sigma_{\substack{\alpha:\, x_{\alpha o}=x_{ao}\\ w_{\alpha t}=\bar{w}_{at}}}\left(P(\alpha \mid s_{to})\,\Pi_{\substack{\beta:\,\beta\neq\alpha\\ x_{\beta o}=x_{ao}}} (1-P(\beta \mid s_{to}))\right)}{\Sigma_{\alpha:\, x_{\alpha o}=x_{ao}}\left((P(\alpha \mid s_{to})\,\Pi_{\substack{\beta:\,\beta\neq\alpha\\ x_{\beta o}=x_{ao}}} (1-P(\beta \mid s_{to}))\right)}\right) =$$

$$\Pi_{a\in C}\Pr(\bar{w}_{at} \mid x_{ao}, s_{to}).$$

DISCUSSION: Four points focus our discussion of the theorem. These include first an interpretation of the theorem's assumptions; second an explanation of the condition that $(x_{ao},\ a\in C)$ be distinct; third a look at alternative assumptions yielding the IRU property; and fourth a comment on the IRU property.

(1) Of the three assumptions, (i) and (ii) place restrictions on the process generating choice problems while (iii) places an informational requirement on the observer. The form of the utility function is left arbitrary.

Statistically, assumption (i) asserts that each decision maker's choice set is the realization of a set of independent draws of alternatives across $\mathfrak{a}$. Substantively, the assumption is most easily interpreted if the existence of a conscious set of decision makers TT controlling $\mathfrak{a}$ is assumed. In this context, the assumption requires 'offers' of alternatives to be mutually independent.

For example, if $\mathfrak{a}$ is a set of jobs, assumption (i) prohibits any interdependence among job offers. In particular, both bandwagon effects and collusion among firms to segment the labor market would violate the assumption.[22]

Assumption (ii) limits the diversity of choice set generating processes across T. The assumption does not, however, require that all $t \in T$ face identical processes.

Assumption (iii) is straightforward in interpretation. As a practical matter, this assumption can be difficult to fulfill.

(2) The theorem demonstrates the IRU property only for choice sets whose observed attribute vectors are distinct. To analyze the more general case where $(x_{ao}, a \in C)$ need not be distinct, let D be an ordered subset of C such that $(x_{ao}, a \in D)$ includes, without duplication, all the distinct elements of $(x_{ao}, a \in C)$. For each $a \in D$, let $n(a)$ be the number of alternatives in C whose observed attribute vector is x_{ao} and designate those alternatives by a_i, $i = 1, \ldots n(a)$. It can be shown that

$$\Pr(\bar{W}_{ct} \mid r_{co}, s_{to}) = \Pi_{a \in D}$$

$$\left(\frac{\Sigma_{\begin{pmatrix} \alpha_i : x_{\alpha_i} \dot{=} x_{ao} \\ w_{\alpha_i t} = w_{a_i t} \end{pmatrix},\ i=1,\ldots,n(a)} \Sigma \left(\Pi_{i=1}^{n(a)} P(\alpha_i \mid s_t) \right) \left(\Pi_{\substack{\beta:\beta \neq \alpha_i,\ i=1,\ldots,n(a) \\ x_{\beta o} = x_{ao}}} (1 - P(\beta \mid s_t)) \right)}{\Sigma_{(\alpha_i : x_{\alpha_i o} \dot{=} x_{ao}),\ i=1,\ldots,n(a)} \Sigma \left(\Pi_{i=1}^{n(a)} P(\alpha_i \mid s_t) \right) \left(\Pi_{\substack{\beta:\beta \neq \alpha_i,\ i=1,\ldots,n(a) \\ x_{\beta o} = x_{ao}}} (1 - P(\beta \mid s_t)) \right)} \right) =$$

$$= \Pi_{a \in D} \Pr(\bar{w}_{a_i t}, i = 1,..,n(a) \mid x_{ao}, n(a), s_{to})$$

That is, alternatives with distinct observed attribute vectors have independently distributed utilities. Alternatives whose observed attributes are identical have interdependent utilities.[23]

(3) The assumptions of the theorem are sufficient to induce the IRU property but are not necessary. For example, the reader may verify that the following assumptions can substitute for assumptions (ii) and (iii).

(ii′) For every $(a, t) \in \mathfrak{a} \times T, P(a \mid t) = P(a \mid s_{to})$.

(iii′) The mappings $a \rightarrow x_a$, $t \rightarrow s_t$ and $(x, s) \rightarrow w(x, s)$ are all one-to-one.

Moreover, we shall later, within the discussion of random coefficients models, describe an IRU model sharing none of the structural assumptions presented thus far.

As a general matter, sets of minimal structural conditions, whether for the basic random utility model developed in Part I or for the special models examined here, have remained beyond my reach. Efforts seeking to better define the range of validity of these models are to be encouraged.

(4) The independent random utilities property, while helpful, has not by itself proved sufficient to render random utility models analytically or computationally tractable. At this date, only the more restrictive independent and identically distributed random utilities models have found practical application. An examination of this class of models follows.

B. *The Independent and Identically Distributed Random Utilities Models*

Write the utility function $w(x, s)$ as $w(x, s) = V(x_o, s_o) + \epsilon(x, s)$ where $V(x_o, s_o)$ is that part of $w(x, s)$ additively separable in the observed x_o and s_o attributes. Note that for every $a \in \mathfrak{a}, t \in T$, $V_{at} = V(x_{ao}, s_{to})$ is known to the observer but $\epsilon_{at} = \epsilon(x_a, s_t)$ is unobserved. A random utility vector W_{ct} is said to have the *independent and identically distributed random utilities* (IIDRU) property if

$$\Pr(\bar{W}_{ct} \mid r_{co}, s_{to}) = \Pi_{a\in C} \Pr(V_{at} + \bar{\epsilon}_{at})$$

where $\bar{\epsilon}_{at}$, $a \in C$ are realizations of a set of independent and identically distributed random variables. The theorem below gives conditions sufficient to ensure that W_{ct} has the IIDRU property.

THEOREM: *Assume that conditions* (i), (ii) *or* (ii′), *and* (iii) *or* (iii′) *stated previously are met. Also assume the following*:

(iv) *There exists a real function* $\delta(x_u)$ *such that* $w(x_a, s_t) = V(x_{ao}, s_{to}) + \delta(x_{au})$, all $a \in \mathfrak{a}$, $t \in T$.

(v) *For all values of* $x_o, \bar{x}_u$, *and* s_o, $\Pr(\bar{x}_u \mid x_o, s_o) = \Pr(\bar{x}_u)$.[24]

Let (C, t) *be a choice problem such that the observed attributes* $(x_{ao}$, $a \in C)$ *are distinct. Then, for all values of*

$$\bar{W}_{ct}, \Pr(\bar{W}_{ct} \mid r_{co}, s_{to}) = \Pi_{a\in C} \Pr(V_{at} + \bar{\epsilon}_{at})$$

where ϵ_{at}, $a \in C$ *are independent and identically distributed.*

Proof: The theorem's assumptions guarantee that W_{ct} has the IRU property. That is,

$$\Pr(\bar{W}_{ct} \mid r_{co}, s_{to}) = \Pi_{a \in C} \Pr(\bar{w}_{at} \mid x_{ao}, s_{to}) =$$

$$\Pi_{a \in C} \Pr(V_{at} + \bar{\epsilon}_{at} \mid x_{ao}, s_{to}),$$

where $\epsilon_{at}, a \in C$ are independent random variables.

Assumptions (iv) and (v) guarantee that for given a and t, ϵ_{at} is neither functionally nor distributionally dependent on x_{ao} or s_{to}.
Hence, $\epsilon_{at}, a \in C$ have observationally identical distributions and

$$\Pr(\bar{W}_{ct} \mid r_{co}, s_{to}) = \Pi_{a \in C} \Pr(V_{at} + \bar{\epsilon}_{at}),$$

where $\epsilon_{at}, a \in C$ are i.i.d.

DISCUSSION: Our discussion first interprets the two new assumptions introduced by the theorem. We then examine an interesting weakened form of the IIDRU property.

(1) Both assumptions (iv) and (v) place conditions on the 'fine structure' of the choice process.

Assumption (iv) concerns both the form of the utility function and the extent of the observer's information. What seems most restrictive about this separability assumption is its requirement that the sub-utility function $\delta(x_u)$ not be dependent on t.[25] That is, preferences with respect to the x_u attributes are assumed homogeneous among the decision makers T.

Assumption (v) relates to the choice problem generating process and to the observer's information. In its requirement that x_u be distributed independently of s_o, (v) enforces a certain limited uniformity among the choice set generating processes faced by the members of T. In its requirement that x_u be distributed independently of x_o, the assumption is analogous, although stronger than, the usual linear model condition that excluded variables be uncorrelated with included ones.

Because fulfillment of (v) depends in detailed ways on the structure of the mapping from the alternative space to its attribute space and on the nature of the processes generating choice sets, it is difficult to interpret the assumption substantively. Consequently, it is difficult to provide the empirical researcher with guidance as to when he might reasonably conclude that the assumption is satisfied. Lack of any meaningful interpretation of assumption (v) constitutes an important gap in our understanding of the structure of random utility models.[26]

(2) Consider a choice process satisfying the following assumptions:

(i) For every

$$(C, t) \in \Gamma \times T, M_\Gamma(C \mid t) = (\Pi_{a \in C} P(a \mid t)) (\Pi_{\beta \notin C} (1 - P(\beta \mid t))).$$

(iv′) There exists a real function $\delta(x_u, s_o)$ such that $w(x_a, s_t) = V(x_{ao}, s_{to}) + \delta(x_{au}, s_t)$, all $a \in \mathfrak{a}$, $t \in T$.

(v′) For all values of x_o and $\bar{x}_u$, $\Pr(\bar{x}_u \mid x_o, t) = \Pr(\bar{x}_u \mid t)$.

Let C be a choice set such that $(x_{ao}, a \in C)$ are distinct. It is straightforward to show that W_{ct} then satisfies the *conditionally independent and identically distributed random utilities* (CIIDRU) property:

$$\Pr(\bar{W}_{ct} \mid r_{co}, t) = \Pi_{a \in C} \Pr(V_{at} + \bar{\epsilon}_{at} \mid t),$$

where ϵ_{at}, $a \in C$ are, conditional on t, i.i.d.r.v.'s.

The CIIDRU property, avoiding as it does any restriction of the distributional relation between W_{ct} and t, is a significantly weakened form of the IIDRU property.[27] Nevertheless, models possessing this property are amenable to estimation by existing methods. See Manski (1975) where it is proved that under suitable regularity conditions, 'maximum score' estimates of CIIDRU models are consistent.

C. *The Random Coefficients Models*

Consider a choice problem $(C, t) \in \Gamma \times T$. Let $\hat{W}(C, t, \bar{s}_u) = (w(x_a, s_{to}, \bar{s}_u), a \in C)$, all $\bar{s}_u \in S_u$. The random utility vector W_{ct} has the *random coefficients* (RC) property if, for all values of $\bar{W}_{ct}$,

$$\Pr(\bar{W}_{ct} \mid r_{co}, s_{to}) = \Sigma_{\bar{s}_u : \hat{W}(C, t, \bar{s}_u) = \bar{W}_{ct}} \Pr(\bar{s}_u \mid s_{to}),$$

where

$$\Pr(\bar{s}_u \mid s_{to}) = \Sigma_{\tilde{t}:\, \substack{s_{\tilde{t}o} = s_{to} \\ s_{\tilde{t}u} = \bar{s}_u}} M_T(\tilde{t}) \Big/ \Sigma_{\tilde{t}:\, s_{\tilde{t}o} = s_{to}} M_T(\tilde{t}).$$

The theorem below gives conditions sufficient to ensure that a utility vector W_{ct} has the random coefficients property.

THEOREM: *Assume that*

(vi) *For all* $(C, t) \in \Gamma \times T, M_\Gamma(C \mid t) = M_\Gamma(C \mid s_{to})$.

(vii) *For all* $a \in \mathfrak{a}, x_{ao} = x_a$.

Then, for all values of $\bar{W}_{ct}$,

$$\Pr(\bar{W}_{ct} \mid r_{co}, s_{to}) = \Sigma_{\bar{s}_u:\, \hat{W}(C,t,\bar{s}_u) = \bar{W}_{ct}} \Pr(\bar{s}_u \mid s_{to}).$$

Proof: By the assumptions of the theorem,

$$\Pr(\bar{W}_{ct} \mid r_{co}, s_{to}) =$$

$$\Sigma_{\substack{(\tilde{C},\tilde{t}):\, r_{\tilde{c}o} = r_{co} \\ s_{\tilde{t}o} = s_{to} \\ W_{\tilde{c}\tilde{t}} = \bar{W}_{ct}}} M_{\Gamma T}(\tilde{C}, \tilde{t}) \Big/ \Sigma_{\substack{(\tilde{C},\tilde{t}):\, r_{\tilde{c}o} = r_{co} \\ s_{\tilde{t}o} = s_{to}}} M_{\Gamma T}(\tilde{C}, \tilde{t}) =$$

$$\frac{\Sigma_{\tilde{C}:\, r_{\tilde{c}o} = r_{co}} \Sigma_{\substack{\tilde{t}:\, s_{\tilde{t}o} = s_{to} \\ \hat{W}(C,t,s_{\tilde{t}u}) = \bar{W}_{ct}}} M_\Gamma(C \mid s_{to}) M_T(t)}{\Sigma_{\tilde{C}:\, r_{\tilde{c}o} = r_{co}} \Sigma_{\tilde{t}:\, s_{\tilde{t}o} = s_{to}} M_\Gamma(C \mid s_{to}) M_T(\tilde{t})}$$

$$= \Sigma_{\substack{\tilde{t}:\, s_{\tilde{t}o} = s_{to} \\ \hat{W}(C,\tilde{t},s_{\tilde{t}u}) = \bar{W}_{ct}}} M_T(\tilde{t}) \Big/ \Sigma_{\tilde{t}:\, s_{\tilde{t}o} = s_{to}} M_T(\tilde{t}) =$$

$$= \Sigma_{\bar{s}_u:\, \hat{W}(C,t,\bar{s}_u) = \bar{W}_{ct}} \Pr(\bar{s}_u \mid s_{to}).$$

DISCUSSION: (1) In random coefficients models, the existence of unobserved decision maker attributes s_u forms the sole source of observational randomness in W_{ct}. Furthermore, knowledge of the choice set faced by t yields no information as to the value of s_{tu}. In the above theorem, assumption (vi) guarantees the latter property while (vii) guarantees the former. Both assumptions have straightforward interpretations.

(2) The random utility models developed by psychometricians generally fall within the random coefficients class. For example, in Tversky's elimination by aspects model, alernatives have 0–1 'aspects', all of which are observed, and the decision maker draws a lexicographic decision rule from a set of such

rules. The rule drawn in any choice situation is unobserved and is drawn independently of the choice set faced.[28]

(3) Interestingly, there exist random coefficients models which also have the IIDRU property. Assume that $w(x, s) = x_o \cdot s_o + x_o \cdot s_u$ where x_o, s_o and s_u are each K element vectors and s_u is distributed independently of C. Thus, the model has the RC property. Also assume that for every $a \in \mathfrak{a}$, x_{ao} is a permutation of the vector $(1, 0, ..., 0)$ and that the K components of s_u are mutually independent and identically distributed and distributed independently of s_o as well. Letting $V_{at} = x_{ao} \cdot s_{to}$ and $\epsilon_{at} = x_{ao} \cdot s_{tu}$, it follows that if $(x_{ao}, a \in C)$ are distinct, $\epsilon_{at}, a \in C$ are i.i.d.[29]

SOME OBSERVATIONS

My work on this paper was motivated by an uneasiness about the econometric literature's assymetric handling of the 'systematic' and 'disturbance' components of a random utility model. There is an essential symmetry between V and ϵ in the equation $U = V + \epsilon$. Nevertheless, attention in empirical analyses has focussed almost exclusively on the functional form of V to the neglect of the probability distribution of ϵ. Furthermore, theoretical discussion of distributional specification has been very limited. These trends, I felt, needed correction.

At its completion, I have mixed feelings about the success of my endeavor. On the positive side, Part I presents an internally consistent and structurally meaningful derivation of the random utility model. Moreover, as the theorems of Part II indicate, that derivation can be a powerful tool for interpreting the special models found in the literature.

My misgivings, ironically enough, arise out of a perceived assymetry in my work. The classical choice process, Section IC emphasizes, is a recursive one. First, a choice problem is generated and then the specified decision maker selects from his realized choice set. In this paper, a specific rule for the decision maker, namely utility maximization, is assumed but the choice problem generating process is described only as a probability distribution over $\Gamma \times T$. An interpretation of $M_{\Gamma T}$ in terms of the behavior of TT is offered and is helpful but, like the similar-distinct dichotomy, it is not fully satisfying. Thus, this paper may have resolved some questions about the distributional specification of random utility models only to replace them with questions about the distribution governing the generation of choice problems.

Looking beyond the present paper, I see a number of directions for future research, corresponding to the four sets of assumptions laid out in Section IA through ID.

One direction obviously is the structural investigation of choice problem generation. Processes assuming conscious decision makers TT seem particularly ripe for analysis but 'natural' processes should not be ignored.

A second focus would be the study of alternative information bases for the observer. As alluded to earlier, a weakening of our informational assumptions would seem to lead to Bayesian random utility models. On the other side, information not considered in the present analysis is sometimes available to the observer. For example, knowledge of non-utility relevant attributes can be valuable.

Third, the utility maximizing decision rule we assume might be dispensed with or strengthened. For example, Tversky has shown that a conceptually different rule, elimination-by-aspects, is consistent with a random utility model. Lancaster has demonstrated the analytical power that linear attribute spaces yield in the presence of utility maximization.

A final set of questions concerns the structure of the spaces $\mathfrak{a}$ and T. In particular, it is not clear whether the extension of our work to infinite alternative and decision maker spaces involves merely matters of mathematical interest or more.

Carnegie – Mellon University

NOTES

*Assistant Professor of Economics, School of Urban and Public Affairs, Carnegie-Mellon University.

I am grateful to Joseph B. Kadane for numerous constructive suggestions offered during discussions of this research. The financial sponsorship of the U.S. Department of Transportation through grant DOT-OS-4006 is also acknowledged. The opinions and conclusions expressed herein are solely those of the author.

[1] Luce and Suppes (1965) and Tversky (1972b) provide good overviews of this literature.

[2] McFadden (1974) provides a comprehensive survey of this literature.

[3] Debreu's observation was made with respect to Luce's 'strict utility' model before a random utility interpretation of that model had been obtained. Although the problem with the Luce model was first attributed to its peculiar 'independence of irrelevant alternatives' property, it was later recognized by Luce and Suppes (1965) that the strict utility model can be restated as a particular IIDRU model and that all IIDRU models share the Luce Model's deficiency.

[4] The 'similar-distinct' dichotomy arose through interpretation of the examples in which IIDRU models make implausible choice forecasts. In all of those examples, the problem appears when a new alternative, identical to an existing one in all utility-relevant respects, is added to the choice set. The failure of IIDRU models in such a situation is easily shown. If alternatives i and j form the original choice set so that

$$\Pr(i \in^c (i, j)) = \Pr(U_i \geqslant U_j),$$

addition of a third alternative k, identical to j yields

$$\Pr(i \in^c (i, j, k)) = \Pr(U_i \geqslant U_j, U_k) < \Pr(U_i > U_j) = \Pr(i \in^c (i, j)),$$

certainly a counter-intuitive result. The interpretation of the problem to extend to 'similar' as well as identical alternatives seemed straightforward. Nevertheless, a definition of similarity has remained elusive. See Ginsberg (1972), Luce and Suppes (1965), McFadden (1974), (1975), Kohn *et al.* (1974), McLynn and Goodman (1973) and Varian (1974) for discussions of the problem and examples of the ad hoc corrections proposed to resolve it.

[5] Papers dealings with estimation and forecasting methodology include McFadden (1973), McFadden and Reid (1974), Manski (1975), Manski and Lerman (1975), and Westin (1974).

[6] Some aspects of the development are only seemingly restrictive. In particular, although the presentation assigns the source of randomness to incomplete observer information, it does not rule out inconsistent individual behavior. In order to interpret the model in this way, one need only view T as defining not a population of individually rational decision makers but a population of decision rules from which an individual makes random draws.

[7] The utility maximizing model does not specify how the choice among alternatives bearing equal utilities is resolved. For our purposes, it is perhaps easiest to assume simply that each decision maker orders the finite set $\mathfrak{a}$ completely. That is, indifference among alternatives does not exist.

[8] In this case, X may itself be defined as the alternative space.

[9] This becomes relevant when we seek to distinguish between observed and unobserved attributes in Section IID.

[10] Lancaster (1971) assumes linear space properties in his attribute model of consumer demand. His quite restrictive specification does yield analytical results which our weak assumptions cannot.

[11] Formal treatments within the classical framework include Arrow (1958, 1959), Fishburn (1970), Gorman (1968), Luce (1959), and Weddepohl (1970).

[12] For psychological models, it seems more natural to decompose $M_{\Gamma T}$ as follows: $M_{\Gamma T}(\mathcal{C}, t) = M_T(t \mid \mathcal{C}) M_\Gamma(C)$. In this formulation, $M_T(t \mid C)$ is the probability that the decision maker draws decision rule t given that he is faced with choice set C and $M_\Gamma(C)$ is the probability that he is faced with this choice set. We note that the psychological models developed thus far usually do not allow the distribution governing the drawing of choice rules to be influenced by the choice set faced. That is, they assume $M_T(t \mid C) = M_T(t)$.

[13] Because most game theoretic models assume simultaneous decisions on the part of all players, no well-defined concept of a choice set exists for such models. Only equilibrium solutions are well-defined.

[14] An example distinguishing independent from non-independent processes may be

useful. If TT is a set of colleges and every $tt \in TT$ evaluates each of its applicants in isolation from its other applicants, college choice sets will be independently generated. However, if because of capacity constraints, colleges make comparative decisions about applicants and if T contains more than one applicant to a given college, independence is violated.

[15] In these expressions and hereafter, a bar (–) over a variable symbol indicates that it is a realization of an observationally random variable. Thus, $\bar{s}_t$ is some realization of the observationally random decision maker attribute vector while s_t is the unknown true value of this vector.

[16] It might be thought that our model actually allows incomplete knowledge of w, since an unknown parameter of w and an unobserved constant element in x_u or s_u are indistinguishable. However, the assumption that the probabilities $\Pr(\bar{r}_{cu}, \bar{s}_{tu} \mid r_{co}, s_{to})$ are known implies that the value of any constant attribute, even if directly unobserved, can be determined. We therefore must keep unknown function parameters and unobserved constant attributes conceptually distinct.

[17] We do not rule out the potential of Bayes procedures. The random utility model literature has been totally confined to classical point estimation of the unknown parameters of w and application of the estimated function $\hat{w}$ in forecasting as if it were w. It would seem more appropriate, although considerably more complicated, to go to the empirical Bayes route; that is, use the output of the parametric estimation process to obtain a distribution for w and treat w as a random function with this distribution in forecasting.

[18] We adopt the convention that if $a \notin C$, then $\Pr_t(a \in^c C) = 0$.

[19] Expected choice set imputation is well-defined only if $\mathfrak{A}$ partitions into N categories (such as auto, bus, taxi), each choice set with positive probability contains one alternative from each category and convex combinations of alternatives are themselves alternatives. Treating choice sets as ordered sets of alternatives, the expected choice set is then $C = \Sigma_{C \in \Gamma} C M_\Gamma (C \mid t, a \in C)$. It should be pointed out that practitioners have not attempted to define the expected choice set procedure formally. I believe however that the above definition does describe the essence of their practice.

[20] Note that the definition of the IRU property is phrased directly in terms of our development of the random utility model and not more abstractly. More general definitions of the property tend to be ambiguous. For example, Luce and Suppes simply state that an IRU model is one in which the utilities U_{at}, $a \in C$ are independently distributed. Such a definition leaves unclear the information conditioning the utility vector's distribution. See Luce and Suppes (1965) p. 338.

[21] This last assumption ensures that the realized choice set is non-empty.

[22] As a second example, let $\mathfrak{A}$ be a set of transportation mode-route alternatives and TT be a metropolitan transportation planning agency whose 'master plan' is reflected in the existing network. The agency might well view the construction of a new highway in an area of good transit service (or vice versa) as redundant. If so, the probability that a community with given characteristics has both good auto and transit links will be smaller than the product of the probabilities of good auto and good transit. Hence, the independence assumption would be violated.

[23] The interdependence among the utilities of alternatives with identical observed attributes arises because for any x_o, the population of alternatives α: $x_{\alpha o} = x_o$ is finite and the process of drawing alternatives with attributes x_o into the choice set is one of sampling without replacement from that population.

[24] In words, let x_o be the observed attributes of an alternative within the choice set faced by a decision maker with observed attributes s_o. The probability that this alternative has unobserved attributes $\bar{x}_u$ is independent of both x_o and s_o.
[25] Note that in the degenerate case where all alternative attributes are observed, this restriction is not effective.
[26] As an example of the problems in interpreting (v), let $\mathfrak{a}$ be a set of differentiated jobs. Then fulfillment of (v) depends on the structure of the distribution of job offers. Clearly, this distribution itself is determined by the operation of the labor market. It would therefore seem that meaningful interpretation of (v) requires a detailed model of a labor market with heterogeneous jobs and workers.
[27] Note that a CIIDRU model need not even have the IRU property. Such a model is, however, a 'conditionally independent random utilities' (CIRU) model.
[28] See Tversky (1972). Note that Tversky's model is a random utility model in spite of the fact that his decision rules are lexicographic. This follows because his alternative space is assumed finite, implying that a utility function can explain any ordering of alternatives. It seems likely that Tversky's model in the presence of an uncountable alternative space would no longer be a random utility model.
[29] A similar derivation of the IIDRU property is given by Tversky for his elimination-by-aspects model.

BIBLIOGRAPHY

Arrow, K., 'Utilities, Attitudes, Choices: A Review Note', *Econometrica* **26** (1958), 1–23.

Arrow, K., 'Rational Choice Functions and Orderings', *Economica* **26** (1959), 121–127.

Ben-Akiva, M. and Lerman, S., 'A Disaggregate Behavioral Model of Automobile Ownership', presented at the Annual Meeting of the Transportation Research Board, January 1975.

Block, H. and Marschak, J., 'Random Orderings and Stochastic Theories of Response', in Olkin (ed.), *Contributions to Probability and Statistics*, Stanford University Press, 1960.

Debreu, G., 'Review of R. Luce, Individual Choice Behavior', *American Economic Review* (1960), 186–188.

Domenchic, T. and McFadden, D., *Urban Travel Demand: A Behavioral Analysis*, North Holland, 1975.

Fishburn, P., *Utility Theory for Decision Making*, Wiley and Son, 1970.

Ginsberg, R., 'Incorporating Causal Structure and Exogenous Information With Probabilistic Models: With Special Reference to Choice, Gravity, Migration, and Markov Chains', *Journal of Mathematical Sociology* **2** (1972), 83–103.

Gorman, T., 'The Structure of Utility Functions', *Review of Economic Studies* **35** (1968), 367–390.

Kelley, J., *General Topology*, Van Nostrand, 1955.

Kohn, M., Manski, C., and Mundel, D., 'An Empirical Investigation of Factors Which Influence College Going Behavior', Rand Corporation Report R-1470-NSF, September, 1974.

Lancaster, K., *Consumer Demand: A New Approach*, Columbia University Press, 1971.

Luce, R., *Individual Choice Behavior*, Wiley, 1959.

Luce, R. and Suppes, P., 'Preference, Utility and Subjective Probability', in Luce, Bush and Galanter (eds.), *Handbook of Mathematical Psychology III*, Wiley, pp. 249–410.

McFadden, D., 'The Revealed Preferences of a Government Bureaucracy', forthcoming, *Bell Journal of Economics and Management Science* (1968).

McFadden, D., 'Conditional Logit Analysis of Qualitative Choice Behaviour', in Zarembka (ed.), *Frontiers in Econometrics*, Academic Press, 1973.

McFadden, D., 'Quantal Choice Analysis', 1974, unpublished.

McFadden, D. and Reid, F., 'Aggregate Travel Demand Forecasting from Disaggregated Behavioral Models', presented at the Annual Meeting of the Transportation Research Board, January, 1975.

McFadden, D., 'On Independence, Structure and Simultaneity in Transportation Demand Analysis', May, 1975, unpublished.

McLynn, J. and Goodman, K., 'Mode Choice and the Shirley Highway Experiment', Urban Mass Transportation Administration Report UMTA-IT-06-0024-73-1, November, 1973.

Manski, C., 'Maximum Score Estimation of the Stochastic Utility Model of Choice', *Journal of Econometrics* 3 (1975), 205–228.

Manski, C. and Lerman, S., 'The Estimation of Choice Probabilities from Choice Based Samples', October, 1975, unpublished.

Quandt, R. and Young, K., 'Cross Sectional Travel Demand Models: Estimation and Tests', *Journal of Regional Science* 9 (1969), 201–214.

Tversky, A., 'Elimination by Aspects: A Theory of Choice', *Psychological Review* 79 (1972), 281–299.

Tversky, A., 'Choice by Elimination', *Journal of Mathematical Psychology* 9 (1972), 341–367.

Varian, H., 'A Note on Conditional Logit Models and the Independence of Irrelevant Alternatives', 1974, unpublished.

Weddepohl, H., *Axiomatic Choice Models*, Rotterdam University Press, 1970.

Westin, R., 'Predictions From Binary Choice Models', *Journal of Econometrics* 2 (1974) 1–16.