



Dissecting the Random Component of Utility

JORDAN LOUVIERE AND DEBORAH STREET
University of Technology, Sydney

RICHARD CARSON
University of California, San Diego

ANDREW AINSLIE AND J. R. DESHAZO
University of California, Los Angeles

TRUDY CAMERON
University of Oregon

DAVID HENSHER
University of Sydney

ROBERT KOHN
Australian Graduate School of Management

TONY MARLEY
McGill University

February 2002

Abstract

We illustrate and discuss several general issues associated with the random component of utility, or more generally “unobserved variability”. We posit a general conceptual framework that suggests a variance components view as an appropriate structure for unobserved variability. This framework suggests that “unobserved heterogeneity” is only one component of unobserved variability; hence, a more general view is required. We review a considerable amount of empirical research that suggests that random components are unlikely to be independent of systematic components, and random component variances are unlikely to be constant between or within individuals, time periods, locations, etc. We also review evidence that random components are functions of (elements of) systematic components. The latter suggests considerable caution in the use and interpretation of complex choice model specifications, in particular recently introduced forms of random parameter models that purport to estimate distributions of preference parameters. Several areas for future research are identified and discussed.

Key words: choice models, random components, unobserved variability, heterogeneity

A Conceptual Framework for Understanding the Issues

The vast majority of research in economics, psychology, social science and business focuses on the effects of manipulated or observational variables on the mean of one or more outcome distributions. Rarely does one see researchers express interest in the effects

of research variables on the variance of outcome distributions. (There are, of course, exceptions including research in risky decision-making, some areas in finance.) The purpose of this paper is to summarize what we know about the behavior of the variance of outcome or response distributions, how effects on response variability can influence empirical outcomes and what can be done about it, and some useful directions for future research.

Our attention is focused on the variance of the random component of utility, which can be thought of as “unobserved variability”. Specifically, we look at unobserved variability in choice models as a behavioral phenomenon, what we know about it empirically, the role that it plays in observed outcome distributions and, fruitful areas of future research that can expand our understanding of this phenomenon.¹ However, let us first consider why researchers should be interested in and care about unobserved variability. We suggest that there are at least five reasons why:

1. *Developing more general models of choice or demand.* To develop meta-models of choice behavior that will be robust to a variety of conditions one should identify and control for sources of variability.
2. *Improving estimates of the parameters of systematic components.* To obtain more reliable and valid estimates of parameters of systematic components one should identify and control for random component variability.
3. *Combining Data Sources.* To facilitate combining data sources for purposes of generalizing systematic analysis into new domains, expanding statistical power with larger samples or testing the reliability and validity of SP methods (or vice versa) one should identify and control for random component variability.
4. *Forecasting or Simulating Choice Behavior.* To improve our ability to forecast choice behavior by characterizing both systematic and random components more completely one should identify and control for variability in the random component.
5. *Testing theories (accounts) of the choice process.* To support innovative ways to evaluate theories of rational choice. Theories that predict differences in variability due to changes in context and time may provide rival explanations to other alternatives to rational choice theory. For example, de Palma, et al. (1994) posit an alternative to standard rational choice that can be tested by analysis of the random component and DeShazo and Fermo (2001b) provide an empirical test of a rival to standard rational choice.

These reasons require further discussion. If a researcher’s main goal is to identify and estimate the parameters of the indirect utility function in a probabilistic discrete choice model, the implicit variance of the random component of utility can be viewed as a nuisance parameter. That is, if there are differences in random component variances across contexts in which choices are observed, failure to allow for these differences can distort utility estimates. In practice, most analysts seek to estimate a demand function (or inverse demand function) that can be transferred to specific contexts and consumers whose choices have not yet been observed. Hence, one typically wants to predict the demand for individuals who face certain vectors of prices, have certain incomes and can be characterized by observable characteristics that act as proxies for systematic preference

variations. One cannot predict the exact quantity demanded, but one can estimate the distribution of the unknown variable, specifically a point estimate and an associated interval estimate.

For example, consider a binary discrete choice model where one implicitly assumes that the latent individual utility difference between two options depends on some function of observable variables and an additive logistically distributed random component. Since the location and scale of the utility difference are unobservable, one can standardize the conditional variance of the utility difference to unity and adjust the location so that the choice of one option implies a positive utility difference, and choice of the other implies a negative utility difference. To make the example specific, consider a situation where a direct mail offer to buy a particular product where different prices were randomly assigned to different offer recipients and a simple probit model is fit to the data with the dependent variable being whether the offer was accepted. In this example, the standard normalization sets the implicit utility difference random component variance to one, and the random component variance is functionally related to the size of the estimate of the effect of the price offered in the model. In particular, the mean price at which there would be a 50% acceptance rate is given by the intercept (location) parameter given divided by (minus) the slope parameter on the price offered (Cameron and James, 1987). The offer price acceptance (i.e., the willingness to pay (WTP) distribution) are completely defined by the intercept and slope parameters in the probit model, with the slope parameter being equal to the negative inverse of the standard deviation of the willingness to pay distribution. Thus, a zero value for the slope parameter implies there is no dispersion in the sample's willingness to pay for the product.

If we were to make the same set of price offers but this time using a telephone solicitation, it might not be surprising to have a different fraction of the sample accept the offer made at any particular price than to the comparable direct mail offer. It may well be that the population of interest's utility function for the good has not changed but rather the "noise" level with respect to making choices differs between the direct mail and telephone solicitation environments. Now consider the implications of pooling the data from these two experiments. If the utility-difference random components systematically differ across the two samples and the utility parameters are constrained to be identical, then the point estimates of the parameters of the WTP distributions from both samples must be identical (and likely wrong). In contrast, if one allows for a possible difference in the utility-difference random component variances (via estimating separate price parameters for the mail and telephone samples), this will propagate through the calculation of WTP distributions from the fitted utility-difference function. If the random component variances differ, then the fitted WTP distributions for each sub-sample will have different variances but the same mean.

If one disallows negative WTP values, as is plausible for many consumer products, then predicted point estimates of WTP in linear models will not be independent of random component variances for the estimated WTP distribution. For example, if negative WTP values are precluded by using an indirect utility function that is log-linear in net income, setting the utility-difference equal to zero and solving yields an expression for predicted WTP that involves the exponentiation of the additive utility-difference error term. As with

simpler log-linear regression models, exponentiating only the systematic part of the model produces a biased estimate of the expected value of the distribution of the desired variable. Thus, random component variances in random utility theory-based choice models can be highly consequential for demand forecasting, even though they may seem to be only “nuisance” parameters.

For example, many practitioners calculate welfare estimates from fitted random utility models using only the systematic component of the fitted indirect utility function, and pay no attention to the random component after estimation, which is of little concern if random components are drawn from an identical distribution for every observed choice. However, utility parameters also can be constrained to be equal across sub-samples, which allows different random component variances by normalizing the variance for one subsample and estimating variances for other subsamples relative to that subsample. In the latter case, different random component variances can lead to different predicted distributions of welfare estimates, even if one constrains the systematic component parameters to be equal for sub-samples.

Having noted, illustrated and discussed reasons why researchers should care about unobserved variability, we now consider the role of unobserved variability in the context of a more general conceptual framework for understanding behavioral outcomes of academic and practical interest. Let us consider a vector of behavioral outcomes of academic and/or commercial interest that we denote Y , such as discrete choices, purchase volumes, profits, or attitudes. In any one empirical study, such outcomes are conditional on a variety of types of dimensions: i) directly manipulable or observable variables that we term X , such as prices, ads, product features, etc; ii) characteristics of individuals whose behavior(s) are observed that we term Z , such as age, income, or psychographics; iii) conditions, contexts, circumstances or situations that we denote as C , such as different trip purposes, types of purchase occasions, or complete/incomplete information conditions in experiments; iv) geographical, spatial or environmental, characteristics that are relatively constant in one place but may vary from place to place that we call G , such as travel times, climate, distribution channels, channel coverage, etc; and v) particular time slices or periods that we term T . Thus, we can write outcomes as conditional on these types of variables as follows:

$$Y|X, Z, C, G, T, \tag{1}$$

where all terms were defined immediately above.

Each of these vectors is multidimensional in the sense that each can represent an array of variables, and moreover each variable can take on a range of values (quantitative or qualitative). Thus, empirical researchers face a highly complex, potentially very high-dimensional combinatorial problem. For example, X spans a large set of variables, say $\{x_1, x_2, \dots, x_X\}$, and each variable, such as price, can take on many levels; Z spans a set of variables, say $\{z_1, z_2, \dots, z_Z\}$, and each variable, such as age, can take on many levels, and similarly for $C \{c_1, \dots, c_C\}$, $G \{g_1, \dots, g_G\}$, and $T \{t_1, \dots, t_T\}$. Each variable can take on values or levels (possibly infinite for continuous variables, perhaps only a few for qualitative variables). Thus, the total possible combinations of X , Z , C , G and T are given by the factorial of the levels of the variables that comprise each, and while the

number of combinations is finite, it can be practically infinite. Note that it is possible for each of these factors to influence both variance components and differences in means. We see as a key research question whether modelling attention should now be focused on modelling the variance components rather than ever increasingly complex specifications of the mean component when that complexity is driven by combining observations over X , Z , C , G or T dimensions.

Unfortunately, much research in the social and economic disciplines is characterized by one-off, non-replicated studies that constitute one combination from this vast factorial array. In fact, viewed in this way, one can see why there has been little progress in many social and economic sciences. That is, there is little agreement on what to measure (i.e., the X 's, Z 's, etc), much less how; most X , Z , C , G , T variables are constant in any one study; and little attention has been given to generalizing from any one empirical research effort. Thus, as Louviere, et al. (1999) noted, it not only is unclear how to generalize from a single empirical study, but most research also is not replicated, which suggests a serious need for research on multiple empirical instances to infer whether phenomena generalize. Thus, the question is not if one should combine data sources to advance understanding, the issue is why isn't it obvious that it must be done? Similarly, generalizations would be more readily accomplished if reviewers and editors were more willing to encourage appropriately structured replications.

Eq. (1) also allows a formal statement of the overarching issue in this paper that deals with the behavior of the variance of the random component. Let us note the basic axiom in random utility theory (RUT):

$$U_{in} = V_{in} + \varepsilon_{in}, \quad (2)$$

where, U_{in} is the unobservable or latent utility that individual n associates with choice option i ; V_{in} is the systematic, observable (explainable) component of utility that can be estimated from behavioral data assuming the data are consistent with RUT; and ε_{in} is the random, unobservable (unexplainable) component of utility that represents researchers' inability to ever fully observe or understand all facets of behavior germane to particular behavioral outcomes of interest. This paper is concerned with $\text{var}(\varepsilon_{in})$ and how it can be decomposed and understood in a behavioral sense.

Returning to Eq. (1), it seems reasonable that $\text{var}(\varepsilon_{in})$ can be generally decomposed into components associated with each of the right-hand side terms in Eq. (1). That is, at the finest level of generality, each variable associated with each construct on the right-hand side of (1) can impact the $\text{var}(\varepsilon_{in})$ in Eq. (2). In addition, there is a level of irreducible response variability (or error) inherent to human behavior that we cannot explain regardless of how complete the right-hand side specification of Eq. (1) is. So, instead of a single unidimensional error component, it would seem more appropriate to posit components of variance expression such that response variability is partitioned into variability due to one or more X 's, Z 's, C 's, G 's and T 's, plus a level of irreducible variability that does not depend on any of these variables. Such a general formulation allows several conclusions about research in the social and economic sciences:

1. Generalizability should be a much more important criterion of empirical research than typically is or has been the case; and, a good starting point would be the slowly emerging literature on Generalizability Theory (e.g., Cronbach et al. 1972; Shavelson and Webb 1991; Finn and Kayande 1997, 1998).
2. Replication is essential for progress and to determine the degree of generality of particular empirical research efforts, not to mention to determine whether other researchers can obtain the same outcomes under identical empirical conditions.
3. It probably is not possible to avoid consideration of combining multiple data sources in order to achieve insights and permit generalizations to be made.
4. Variability in response outcomes and random components can be associated with many factors, and should not simply be lumped either into a single “error term” or attributed entirely to heterogeneity.
5. The variability of the random component of response outcomes is as much a behavioral phenomenon as the variability in mean response outcomes that is the focus of the vast majority of empirical research.

Points 4 and 5 deserve some elaboration. What many researchers call “unobserved heterogeneity” might be better termed “unobserved variability” because Eq. (1) should make it obvious that there are many potential sources of unobserved variability, of which differences in humans/consumers is only one. Thus, research would benefit from a significant switch in focus away from heterogeneity and towards *all* relevant sources of unobserved variability. Of course, in data sources involving consumers, one naturally tends to think that individual differences explain differences in behavioral response outcomes. However, Eq. (1) suggests that this is only one aspect of unobserved variability, hence it is likely that heterogeneity observed in any one data source is conditional on other sources of variability on the right-hand side of Eq. (1). Therefore, it is unclear how to use models estimated from what is effectively a sample of size of one of the possible distributions of consumer heterogeneity in one data source to generalize to other instances, regardless of how well one captures that one source statistically.

Put another way, despite great progress in developing ever more powerful and complex models that can capture virtually all aspects of choice behavior, it nonetheless is the case that such models are only as good as the data from which they are estimated. Eq. (1) makes it clear that many results are potentially context-dependent in so far as behavioral outcomes depend not only on X 's and Z 's, but also on particular factorial combinations of C , G and T in which they are embedded. Failure to take these into account, as well as failure to take Z 's into account properly in complex models calls generalizability into question. This suggests the need to give serious consideration to likely interpretations of effects that are observed/captured/modelled in complex statistical models like mixed logit, random parameters hierarchical Bayes specifications and the like (e.g., Ainslie and Rossi 1998; Allenby and Rossi 1999; Desarbo, et al. 1997; Shively, Allenby and Kohn 2000; Revelt and Train 1998; Train 1997).

For example, little attention has been paid to confounds between distributions of $\text{var}(\varepsilon_{in})$ (the random component associated with each choice option for individual n) and distributions of utility parameters (β s) estimated from choices of samples of consumers

in so-called “random parameter” specifications. That is, it should be obvious that each individual in a population of consumers potentially has a unique systematic and random component. Thus, instead of a single “error” component associated with each individual with certain assumed properties (e.g., IID extreme value type I), it makes more sense behaviorally to assume that each individual has both unique systematic and random components, implying distributions of both systematic and random components, and hence, a distribution of $\text{var}(\varepsilon_{in})$.

As noted by Swait and Louviere (1993), and discussed by others (e.g., Louviere et al. 1999; Hensher, Louviere and Swait 1999), the scale of parameter vectors in all discrete choice models is inversely proportional to the standard deviation of the random component. Thus, if there is a distribution of $\text{var}(\varepsilon_{in})$ across a population, it also follows that each person has a unique scale factor that is inversely proportional to their $\text{var}(\varepsilon_{in})$.

In turn, this implies that each person has a scale factor that is perfectly correlated with their β vector; hence the distribution that is estimated in random parameter discrete choice models is indeterminate. A determination is needed to see if it is a distribution of scale factors, a distribution of β s, or a mixture of both. Moreover, it is unclear how to separate these effects without extra information; hence a fundamental identification problem exists that has statistical and behavioural aspects. Statistically, we need to understand the conditions under which identification can be achieved and the implications of identification issues for empirical work. Behaviourally, empirical research reviewed later suggests that random components vary systematically with respect to a variety of factors under researcher control and/or that can be measured and taken into account (e.g., Cameron, et al. forthcoming; Dellaert, Brazell and Louviere 1999; Swait and Adamowicz 2001a,b). This brings us logically to a review of previous research that bears on the issue of the behavior of the random component.

How we Know that Unobserved Variability Matters

Increasing interest and research into unobserved variability (or the behavior of the random component) began with the pioneering work of Ben-Akiva and Morikawa (1990), who noted that RUT predicts a particular relationship between preferences or behavior observed in stated preference experiments/surveys and behavior/preferences in real markets. That is, they deduced that the parameters of choice models estimated from stated and observed behavior should be proportional if the same utility function underlay both data sources, and the constant of proportionality should be equal to the ratio of the respective $\text{var}(\varepsilon_{in})$ for each data source. This result encouraged others to examine similar relationships between various sources of data consistent with random utility theory. Several early research efforts were cited by Louviere, et al. (1999), with more recent literature reviewed in Hensher, Louviere and Swait (1999) and Louviere, Hensher and Swait (2000, Chapters 8 and 13).

Indeed, as McFadden (2001) noted in his 30-Year Retrospective on Choice Modelling, as well as in several seminars in the US, Australia and elsewhere in 2000 and 2001, it is now well-documented that choice model parameters estimated from revealed and stated sources of preference data typically provide similar conclusions about preferences when stated

preference studies are properly designed and executed. Thus, it is remarkable and important that despite the issues that Eq. (1) poses for generality of process and results, parameters estimated on the X components of Eq. (1) have proved to be very similar over a large number of empirical studies encompassing a wide range of variation in Z 's, C 's, G 's and T 's.

Another important random component issue is the question of why and how one might want to model systematic or random components or both components, as suggested in initial sections of this report. Eq. (1) provides an answer for the "why", which is that in any empirical modelling effort many potential sources of variability implicit in Eq. (1) are held constant. Ideally, one should try to include/model as much systematic behavior as possible, but many factors limit one's ability to do this in a way that can be summarised as data limitations and/or deficiencies. That is, as noted by Louviere, et al. (1999), to better understand/model systematic sources of variation, one needs to combine sources of preference information to estimate effects in Eq. (1) that are constant in some data sources but not others. Absent that, a next best recourse is to try to capture as much of what remains in the random component. The latter has been the focus of research efforts for at least two decades, with considerable activity in the past decade culminating in new forms of random parameter models that permit very general specifications to be estimated. In fact, as shown by McFadden and Train (2000), mixed logit specifications now permit one to approximate all choice processes to an arbitrary degree of accuracy, and the widespread availability of and access to high-speed computing resources now allows many researchers to do this.

However, despite great advancements in theory and methods, issues and limitations associated with trying to model random components remain. One obvious limitation is that one trades-off specific, observational information in systematic components for "latent" information in random components. That is, one faces different limits in each real data source to inferring unobserved influences. The research community has paid relatively little attention to the adequacy of empirical data sources in terms of numbers and quality of observations to support estimation of various types of models, although it may be that Bayesian estimation methods may be less sensitive to sample sizes than classical asymptotic theory methods. Thus, it may well be (and likely is) the case that very many more observations are needed to estimate reliable models for problems of much complexity or dimensionality. So far, most published applications of complex models of random components have used relatively small datasets, and multiple data sources rarely were integrated (although see Ainslie and Rossi (1998) for an example of using multiple data sources to try to understand variability in behaviour).

As earlier noted, it makes behavioral sense to think of a population of consumers that is characterized by a distribution of both systematic and random components, but there is little evidence in recent work on complex model specifications that much attention has been paid to the confound of scale and model parameters. Furthermore, Eq. (1) suggests that latent distributions are likely to be highly context-dependent in the sense that with many effects held constant in single data sources, it is unlikely that estimates of such latent distributions will generalize well. In turn, this suggests that much more research attention should be directed to modelling systematic components and/or expressing the parameters of latent distributions as functions of observables that can be generalized across data combinations in Eq. (1). Indeed, the focus of much of the empirical research that treats the

random component as a behavioural phenomenon is in parameterizing the random component as a function of observables.

The latter issues logically bring us to review some of the conceptual and empirical work in this research stream. Much research work in this rapidly growing paradigm can be described as demonstrating that random component variance can be systematically manipulated under experimental conditions and/or that random component variability differences can be expressed as functions of observables hypothesized to drive it or be associated with it. Ideally one would like to make statements about the conditions under which the random component is likely to vary with X , Z , C , G , T in different contexts but we are not yet at that point. Instead, experiments in this area are in their infancy and have largely served to demonstrate that the random component can be systematically manipulated. We now turn to a brief description of some of these experiments.

Experiments Designed to Systematically Manipulate the Random Component

Brazell, Gray-Lee and Louviere (1998) examined the effects of survey length and scenario order on parameters of the systematic component and unobserved variability in two empirical studies involving the following conditions on the numbers of choice sets: 1) 12, 24, 48, 96 choice sets, and 2) 16, 32, 64, 128 choice sets. They randomly assigned subjects to one of these conditions, and tested if attribute utilities for each condition were equal after controlling for differences in error variances using a FIML extension of the test proposed by Swait and Louviere (1993). That is, they tested if $\beta_1 = \beta_2 = \dots = \beta_L$, after allowing variance scale ratios to differ between the length conditions. In the first study, holiday activity packages were described by orthogonal combinations of ten 2-level attributes that were combined with their foldover pairs to make 12 pairs (+ “neither”). Pairs for 24, 48 and 96 choice sets were made by permuting columns of the 12 choice set design. They found that model parameters estimated from the four length conditions did not differ once differences in random component variability were taken into account. Study two was similar, with identical results. An additional feature of both studies was that the order of the pairs in the smallest choice set conditions (16 and 12 choice sets, respectively) were varied using a Latin square to test order effects on model parameters and unobserved variability. Model parameters were equal for all orders once unobserved variability was taken into account, and there was no relationship between order and relative levels of unobserved variability. The latter finding contradicts notions that subjects use “early” scenarios to learn tasks such that decision strategies change across order. Both results also contradict notions that subjects become tired, confused, bored, etc, as task length increases. No systematic relationships were found between survey length or scenario order and unobserved variability.

Bunch et al. (1998) investigated the effects of survey medium, number of choice options and number of attributes by manipulating the following: 1) paper and pencil versus computer survey administration, 2) 3 versus 6 choice options and 3) 6 versus 12 attributes for two product categories (canned soups & toothbrushes). They used a 2^{4-1} orthogonal fraction to design eight experimental conditions, and randomly assigned 32 subjects to

each of the eight conditions. All subjects received 16 choice sets from an L^{MA} design (Louviere and Woodworth 1983): designs for each condition were based on a common orthogonal design (L = levels, M = choice options and A = attributes). Subjects had to scroll in computer tasks, and could not change earlier answers or look ahead. Results showed that random component variance differed significantly only for five choice options and 12 attributes for canned soups; this difference was roughly three times larger than other conditions. Alternative-specific constants (model intercepts) also differed significantly between computer and paper; hence one would need to combine experimental with real market data to predict accurately in real markets.

Yeoh et al. (1998) investigated the effects of missing attribute information on systematic and random components. They varied the following factors: a) branded (e.g., Levi's, Ford) versus generic options (e.g., A, B); b) numbers of attributes (3 or 6); c) yes/no responses versus trinary choices; and d) two product categories (orange juice versus painkillers). They randomly assigned 200 students to each of the eight conditions, and each subject evaluated 32 scenarios. A FIML extension of the Swait and Louviere (1993) approach was used to test for model parameter equality after taking unobserved variability into account, which indicated that model parameters were not affected by experimental conditions, but random component variances and alternative-specific constants were systematically affected by the conditions. In the random component case, main effects of (more) choice options, (branded) options and (more) attributes increased random component variance, and most interactions also were significant, such that more choices and more attributes significantly increased unobserved variability.

Dellaert, Brazzell and Louviere (1999) studied the effects of varying absolute price levels and price level differences on choices among package tours for three price level differences conditions: equal prices, \$15 price difference, \$45 price difference. The equal price levels were \$40 and \$55, the \$15 price difference levels were \$40 and \$55 and the \$45 price difference levels were \$25 and \$70. In each price level difference condition absolute tour prices varied over two of the levels (\$25, \$40, \$55, \$70). Other attributes varied were length (1/2, 1 day), destination (urban, rural), travel time (1/2, 1 hour) and lunch included (yes, no). A 2^{5-2} design was combined with its foldover to make eight pairs in each price level difference condition (=24 total pairs). Two hundred and four consumers were randomly assigned to the price level conditions and offered eight choices of two bus tours or not going. DBL found that unobserved variability decreased as absolute price levels and price level differences increased. All attribute effects were significant (linear and quadratic price main effects). The systematic component quadratic price effect was significant, for as prices increased, tour choices decreased; but unobserved choice variability also increased, resulting in more choice heterogeneity. The latter effect suggests that as prices increased, price sensitivity heterogeneity increased.

Ohler et al. (1999) studied how attribute range differences impacted binary response outcomes and random component variability in a conjoint analysis task by varying attribute ranges over two levels (a wide range, and middle levels in the wide range). They used a 2^3 design to vary the ranges (wide, middle) of bus fare, time and service, which created eight range conditions, to which they added a ninth condition to test for response non-linearity (a 2^3 factorial was used in each condition to describe buses).

Results showed that attribute ranges did not affect systematic component parameters, but systematically affected attribute interactions and model fits. Moreover, systematic differences in random component variability were found to be associated with omitted attribute interactions; model intercepts also differed systematically with attribute ranges. The latter finding suggests that experimental data must be combined with real market data to predict real market choices accurately. These results suggest that model parameter distributions depend on attribute ranges, as do inferences about non-linearities and interactions (similar to McClelland and Judd (1993) for continuous dependent variables).

Swait and Adamowicz (2001a) investigated the effects of choice set and task complexity on (inter alia) random component variability. They developed a theoretical model that takes into account task complexity, decision-maker effort and ability to choose and choice. They proposed a measure of task complexity, which was incorporated into analyses of several sources of choice data to test whether task complexity was associated with increased variability in factors that affect decision context. They showed that task complexity affects inferences about choice model parameters, and that context effects like complexity systematically impact the parameters of choice models. Specifically, they found an inverted-U relationship between variance and complexity, such that, unobserved variability increases up to a point, then decreases, suggesting the existence of optimal design complexity from a respondent perspective. This is consistent with Severin (2001) immediately below.

Severin (2001) investigated relationships between statistical efficiency (i.e., precision of estimated model parameters) and respondent efficiency. She suggested some statistically efficient experiments may increase cognitive complexity, and as complexity increases choices may become less consistent, causing “respondent efficiency” to decrease. Severin varied design efficiency by varying the number of attribute level differences between pairs of options based on results in Street, Bunch and Moore (2001), who showed that statistical efficiency is a function of numbers of attribute level differences (see also Burgess and Street 2001). Severin also proposed that many statistically optimal designs would be optimal in overall efficiency in cases where only a few attributes are varied because the resulting small number of differences should minimally impact respondent efficiency. However, she proposed that for more attributes the statistically optimal designs would lose overall efficiency because respondent efficiency should decrease significantly. To test these suggestions she varied numbers of attribute level differences for 6 and 16 attributes in two product classes. Respondent efficiency decreased with increasing attribute level differences for six attributes, but did not offset statistical efficiency. Respondent efficiency declines were so large for 16 attributes that optimal designs did not offset declines in respondent efficiency. She also found that attribute numbers *per se* do not impact respondent efficiency once numbers of attribute level differences are taken into account. Hence, the number of attribute differences seems to underlie choice variability, not numbers of attributes *per se*.

Preference Elicitation Procedure Effects

Cameron et al. (forthcoming) studied differences in systematic component model parameter and willingness-to-pay (WTP) associated with one actual purchase and six

widely used hypothetical choice (preference elicitation) formats. They randomly assigned subjects to the seven preference measurement conditions and estimated choice models from each condition. They also pooled the data across all conditions to test the hypothesis that model parameters were the same after accounting for differences in unobserved variability. They found differences in estimated WTP for the individual models that were consistent with earlier findings in the environmental and resource economics literature. However, they also found they could not reject the hypothesis that there was a common underlying systematic component (model parameters) for at least four (and possibly five) of the seven elicitation methods.

DeShazo and Fermo (2001a) investigated how design of choice sets in stated preference surveys impacts unobserved variability. They identified five dimensions of choice complexity and used them as a basis for formulating hypotheses and making predictions about their effects on random component variability. They considered two types of complexity measures: a) those that measure amounts of information within choice sets; and b) those that measure correlational structures of information in choice sets. They used empirical data from two stated preference surveys in Central America to analyse respondents' choice consistency by specifying and estimating a heteroscedastic MNL model in which the scale factor was specified to be a function of the five dimensions of complexity then proposed and tested five competing hypotheses, and showed that their five measures of complexity significantly impact random component variability to different extents and with different significance.

DeShazo and Fermo (2001a) studied how individuals deal with information complexity in the context of rational choice theory assumptions that individuals will use all information provided. If individuals do not use all information, valuation results from traditional choice models would be flawed. They show that individuals selectively discard information as choice set complexity increases, making use of only partial information. They propose a model of the process by which individuals rationally adapt to limited cognition, and hypothesize that individuals select more restrictive processing strategies in response to increasing choice set complexity. They test this hypothesis by developing measures of choice set complexity that are randomly assigned as choice experiments in the context of choices among vacation tour packages. They show that one can improve empirical discrete choice models with better specification methods and more refined experiments. They allow that one should try to identify and control for partial information strategies of individuals as well as try to capture impacts on conditional variances (or scale factors) in discrete choice models. Hence, their results suggest that choice set complexity affects the structure of individual choice processes by affecting the type and quantity of information used to make choices.

Other results for preference elicitation procedure comparisons include the following:

1. Swait and Louviere (1993) compared bike trail discrete choice and consideration responses from 367 Chicago bike riders and found that the systematic component model parameters were the same once random component variability was taken into account.
2. Louviere, Fox and Moore (1993) compared choice models from six stated preference elicitation methods for a convenience sample of 158 subjects in Edmonton, Canada,

and found that the systematic component model parameters were the same for five of the elicitation methods once random component variability differences were taken into account.

3. Swait, Louviere and Williams (1994) compared sources of preference data that included three revealed preference or actual choice data sources and three stated preference data sources for choice of freight shippers in three cities. They found that the systematic component model parameters were the same for five of the elicitation methods once random component variability differences were taken into account for five of the six data sources.
4. Severin, Louviere and Finn (2000) compared stated and revealed preference data sources for choices of shopping centers and supermarkets over time and across cities. They found that the systematic component model parameters were the same for all over time periods and across cities conditions once random component variability differences were taken into account.

Discussion and Conclusions

We noted and discussed several reasons why researchers should be interested in the behavior of the random component of utility, which can be categorized as improving and enhancing modelling capabilities and our ability to gain insights into this process. We also proposed and discussed the implications of a preliminary conceptual framework within which to consider the behavior of the random utility component in decision-making and choice behavior. This framework suggests that most research in the social and economic sciences lacks generality, is context-specific and dependent and effectively represents a sample of size one from the possible conditions that could be investigated. To wit, in almost all empirical research published to date, many sub-components of variability that comprise the random components of utility have been held constant due to data limitations or inadequacies. Thus, any research that tries to model latent phenomena in random components, no matter how strong its theoretical basis, is likely to run into empirical problems due to the need for large amounts of data that vary along X , Z , C , G and/or T in such a way as to support generalization to contexts that differ significantly from the one investigated.

The conceptual framework suggests that unobserved variability is likely to be systematically related to a variety of sources described in the general conceptual framework represented by Eq. (1). Thus, the true state of the world and individuals who behave in it is likely to be characterized by components of variation processes, and distributions of random components. That is, we need to know much more about the distributions of parameters that are used in models that specify both systematic and random components. We noted that if individuals have unique systematic and random components, and each of these components has separate distributions, serious identification issues arise in the interpretation of random effects in probabilistic discrete choice models. Thus, the field could benefit from research into identification issues, whether and which extra information

might sustain identification and the properties of model estimates that result from these investigations.

The latter identification issues are important because they pose issues for those who try to interpret outcomes of random parameter models as distributions of utility parameters. This is so because the distribution of utility parameters is confounded with the distribution of random component variability, and we do not yet know how or when they can be separated. However, it is likely that extra data or additional information in the form of observables will be required. Thus, the research community should give these issues some priority not only to illumine the identification issues, but also to provide a consistent way forward for empirical research and practice.

We also noted that evidence for a behavioural as well as statistical interpretation of random component variability is now substantial and has been growing rapidly since the early 1990s. We reviewed a number of recent studies that show that random component behavior can be manipulated experimentally, and we reviewed other studies that show that random component variability differs systematically in response to observables, such as task complexity, numbers of choice options, number of attribute differences and the like. Random component variability can therefore be viewed in both a behavioural and normative way:

1. Behaviorally, the fact that we now know that random component variability can differ between contexts, places, times and even individuals begs the question of why, and clearly points to the need for both conceptual and empirical research to better define and understand as well as predict the differences. This not only suggests that we need research into ways of modelling variance components processes, but also that we may wish to consider whether unobserved variability should be considered as “nuisance” effects or modelled as a behavioral phenomenon in its own right as part of an overall model of the structure of an individual’s choice process.
2. Normatively, we know that random component variability plays a key role in the empirical results one obtains from statistical choice models, and hence it is important for the research community to begin to consider issues of identification and confoundment. The possibility that individuals have unique systematic and random components also needs to be taken into account in ways that differ from current theory and practice. This leads us to suggest that the field could benefit from guidance as to when and how one should model the systematic and random components and in what ways.

Notwithstanding our need to understand and resolve the above issues, there is encouraging and important evidence of systematic behavior and generality in the behavior of the random component. That is, existing empirical evidence strongly suggests that in many cases there are few systematic differences in the parameters of models of the systematic utility component once differences in random component variability are taken into account.

Finally, the conceptual framework and empirical results that we discussed pose interesting theoretical and empirical problems for research in behavioral decision theory and other areas of “behavioral” social and economic sciences that have reported

violations of preference invariance for many years. Our review suggests that many previously published empirical results probably can be accounted for by failure to take differences in random component variability into account, yet others may simply be experimental artefacts that while systematic and reliable, actually have fairly trivial effects on real choices in real markets, but others remain that currently defy explanation by a random component mechanism.

Notes

1. One reviewer pointed out that the systematic modeling of different sources of variability has a long and rich history in the statistical literature on linear models (Rao, 1973; Verbeke and Molenberghs, 1997). We agree and find the lack of a similar interest in choice models somewhat surprising given the well-known result (Yatchew and Griliches, 1984) that in contrast to the linear model, the parameter estimates of probit and logit models are inconsistent in the presence of heteroscedasticity.

References

- Adamowicz, W., J. Louviere, and M. Williams. (1994). "Combining Stated and Revealed Preference Methods for Valuing Environmental Amenities," *Journal of Environmental Economics and Management*, 26, 271–292.
- Ainslie, A., and P. E. Rossi. (1998). "Similarities in Choice Behavior Across Product Categories," *Marketing Science*, 17, 91–106.
- Allenby, G., and P. Rossi. (1999). "Marketing Models of Consumer Heterogeneity," *Journal of Econometrics*, 89(1–2), 57, 78.
- Ben-Akiva, M., and T. Morikawa. (1990). "Estimation of Switching Models from Revealed Preferences and Stated Intentions," *Transportation Research A*, 24A, 485–495.
- Brazell, J., J. Gray-Lee, and J. J. Louviere, (1998). "Revisiting Order Effects: Are They Just Differences in Random Variation?" Paper presented to the Marketing Science Conference, Fontainebleau, France, July.
- Bunch, D. S. et al. (1998). "Comparing the Effects of Survey Medium, Numbers of Choice Options and Numbers of Attributes on Choices: Maybe It's Just Differences In Error Variability?" Paper presented to the INFORMS Marketing Science Conference, Fontainebleau (INSEAD), France, July.
- Burgess, L., and D. J. Street (2001) "Optimal Designs for 2^k Choice Experiments," Research Report, Department of Mathematical Sciences, University of Technology, Sydney.
- Cameron, T. A., and M. D. James. (1987). "Efficient Estimation Method for Closed-Ended Contingent Valuation Surveys," *Review of Economics and Statistics*, 69, 269–276.
- Cameron, T. A. et al. (forthcoming). "Alternative Nonmarket Value-Elicitation Methods: Are the Underlying Preferences the Same?," *Journal of Environmental Economics and Management*.
- Cronbach, L. J. et al. (1972). *The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles*. New York: John Wiley.
- de Palma, André, Gordon Myers, and Yorgos Papageorgiou. (1994). "Rational Choice Under an Imperfect Ability to Choose," *American Economic Review*, 84(4), 419–440.
- Dellaert, Benedict, Jeff D. Brazell, and Jordan, J. Louviere. (1999). "The Effect of Attribute Variation on Consumer Choice Consistency," *Marketing Letters*, 10(2), 139–147.
- Desarbo, W. et al. (1997). "Representing Heterogeneity in Consumer Response Models," *Marketing Letters*, 8, 335–348.
- DeShazo, J. R., and G. Fermo. (2001a). "Designing Choice Sets for Stated Preference Methods: The Effects of Complexity on Choice Consistency," *Journal of Environmental Economics and Management*, forthcoming.

- DeShazo, J. R., and G. Fermo. (2001b). "Rational Choice with Cognitive Limitations: Empirical Support for Partial-information Processing Strategies," Unpublished Working Paper, Department of Economics, UCLA.
- Finn, A., and U. Kayande. (1997). "Reliability Assessment and Optimization of Marketing Measurement," *Journal of Marketing Research*, 34, 262–275.
- Finn, A., and U. Kayande (1998). "The Service Quality Measurement Literature: A Generalizability Perspective," *Advances in Services Marketing and Management*, 7, 97–130.
- Hensher, D. A., J. J. Louviere, and J. Swait. (1999). "Combining Sources of Preference Data," *Journal of Econometrics*, 89, 197–221.
- Louviere, J., and G. Woodworth. (1983). "Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach Based on Aggregate Data," *Journal of Marketing Research*, 20, 350–367.
- Louviere, J. J. et al. (1999). "Combining Sources of Preference Data for Modelling Complex Decision Processes," *Marketing Letters*, 10(3), 187–204.
- Louviere, J., M. Fox, and W. Moore. (1993). "Cross-Task Validity Comparisons of Stated Preference Choice Models," *Marketing Letters*, 4(3), 205–213.
- Louviere, J. J., D. A. Hensher, and J. D. Swait. (2000). *Stated Choice Methods: Analysis and Application*. Cambridge, UK: Cambridge University Press.
- McClelland, G. H., and C. M. Judd. (1993). "The Statistical Difficulties of Detecting Interactions and Moderator effects," *Psychological Bulletin*, 144(2), 376–390.
- McFadden, D. (2001). "Disaggregate Behavioural Travel Demand's RUM Side: A 30-Year Retrospective," in *Travel Behavioural Research: The Leading Edge*, Editor, D. A. Hensher, Amsterdam: Pergamon, 17–64.
- McFadden, D., and K. Train. (2000). "Mixed MNL Models for Discrete Response," *Journal of Applied Econometrics*, 15, 447–470.
- Ohler, T. et al. (2000). "Attribute Range Effects in Binary Response Tasks," *Marketing Letters*, 11(3), 249–260.
- Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*. 2nd edition. New York: John Wiley.
- Revelt, D., and K. Train. (1998). "Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level," *Review of Economics and Statistics*, 80, 1–11.
- Severin, V. (2001). "Comparing Statistical and Respondent Efficiency In Choice Experiments," Unpublished Ph.D. Dissertation, Discipline of Marketing, Faculty of Economics and Business, The University of Sydney (August).
- Severin, V., J. J. Louviere, and A. Finn. (2000). "The Stability of Retail Shopping Choices Over Time and Across Countries," *Journal of Retailing*, 77, 185–201.
- Shavelson, R. J., and N. M. Webb. (1991). *Generalizability Theory: A Primer*. Newbury Park, CA: Sage Publications.
- Shively, T., G. Allenby, and R. Kohn. (2000). "Identifying Latent Relationships in Marketing: Nonparametric Estimation of Binary Latent Regression," *Marketing Science*, 19, 149–162.
- Street, D. J., D. S. Bunch, and B. Moore. (2001). "Optimal Designs for 2^k Paired Comparison Experiments," *Communications in Statistics—Theory and Methods*, 30, 2149–2171.
- Street, D. J., and L. Burgess. (2001). "Optimal and Near-optimal Pairs for the Estimation of Effects in 2-level Choice Experiments," Unpublished Technical Report, School of Mathematical Sciences, University of Technology, Sydney.
- Swait, J., and W. Adamowicz. (2001a). "Choice Complexity and Decision Strategy Selection," *Journal of Consumer Research*, 28, 135–148.
- Swait, J., and W. Adamowicz. (2001b). "Choice Environment, Market Complexity, and Consumer Behavior: A Theoretical and Empirical Approach for Incorporating Decision Complexity into Models of Consumer Choice," *Organizational Behavior and Human Decision Processes*, forthcoming.
- Swait, J. D., and J. J. Louviere. (1993). "The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models," *Journal of Marketing Research*, 30, 305–314.
- Swait, J., J. Louviere, and M. Williams. (1994). "A Sequential Approach to Exploiting the Combined Strengths of SP and RP Data: Application to Freight Shipper Choice," *Transportation*, 21(2), 135–152.
- Train, K. (1997). "Mixed Logit Models for Recreation Demand," in *Valuing the Environment Using Recreation Demand Models*, Editors, C. Kling and J. Herriges, Brookfield, VT: Edward Elgar.
- Tversky, A. (1969). "The Intransitivity of Preferences," *Psychological Review*, 76, 31–48.

- Tversky, A., and D. Kahneman. (1986). "Rational Choice and the Framing of Decisions," *Journal of Business*, 59(4) (part 2), S251–S278.
- Verbeke, G., and G. Molenberghs, editors. (1997). *Linear Mixed Models in Practice*. New York: Springer-Verlag.
- Yatchew, A., and Z. Griliches. (1984). "Specification Error in Probit Models," *Review of Economics and Statistics*, 66, 134–139.
- Yeoh, A. et al. (1998). "Does Missing Information Affect Means or Variances of Error Distributions?" Paper Presented to the INFORMS Marketing Science Conference, INSEAD, Fontainebleau, France, June.