Observing Unobserved Heterogeneity:

Using Process Data to Enhance Choice Models[1]

Eric J Johnson, Bruce G. S. Hardie, Robert J. Meyer, John Walsh

March 2006

[1]This research is supported by the Columbia Center for Electronic Business. We thank seminar participants at Stanford University, Carnegie Mellon and at Intel Corporation, Greg Allenby and Gary Russell for helpful comments.

Abstract

People use different strategies when making choices. Modeling this choice process heterogeneity, however, is difficult using just the data provided by most standard choice experiments. We try to capture process heterogeneity by augmenting choice models with variables derived from information-acquisition data gathered unobtrusively during choice tasks. These variables supplement standard logit specifications which identify how an individual used the attributes and attribute values to screen and rank alternatives in making a choice. The approach improves in-sample fit, prediction in a holdout sample, and residuals indicate that the models are providing better specified estimates of choice probabilities.

Introduction

Understanding consumer choice presents many challenges. In order to predict how markets make choices, we need to understand how the individuals who make up a market differ. One source of variance is *preference heterogeneity,* or how consumers differ in their tastes for product attributes, such as the weight given to an attribute in a trade-off process (Allenby and Rossi 1999; Hutchinson et al. 2000; Kamakura et al. 1996). But consumers can also differ in the rules or policies they use to process such attribute preferences, a source of variance that we term *process heterogeneity.* We know, for example, that while compensatory trade-off policies are often used given small choice sets, non-compensatory processes that use one attribute to eliminate alternatives tend to be used given larger sets (Johnson and Meyer 1984; McClelland et al. 1987; Payne 1976; Swait and Adamowicz 2001).

This paper focuses on the problem of capturing and describing process heterogeneity. We ask whether click-stream data describing how individuals gather information about alternatives in a computer-based choice experiment can improve the predictive and diagnostic value of choice models. We focus on click-stream data as a source of information about processing differences because they can be gathered as a routine part of any computer-mediated choice task (such as might be conducted on the World-Wide Web), and have a long-history of value as a guide to the study of choice processes (e.g., Payne et al. 1993). The methods we develop, however, would directly apply to other process data such as eye fixation recording or verbal reports.

We first provide a background, reviewing the behavioral evidence describing process heterogeneity in choice. We then examine how we might measure difference in choice processes. Next, we propose a method which could be used to extract information about choice processes in choice models in estimation and apply that information in prediction. Finally, we provide an empirical example which demonstrates significant promise.

Process Heterogeneity and Choice

Describing Process Heterogeneity

Choice processes differ across people, contexts and task environments. Process tracing data such as information acquisition (measured by manual retrieval, computer acquisition, head and eye movements), the use of concurrent and retrospective self-reports, and memory measures, all indicate that most choices result from a mélange of decision processes (see Bettman et al. 1991 for a review). However, this raises an interesting question: Is there anything systematic in such observations of process that can increase our ability to understand, explain and forecast choice?

Many of the shifts in strategy described in the behavioral literature on choice are simplifications or heuristics that are used to make choices easier. At one level these choice heuristics can be thought of as complete decision procedures, such as Elimination-by-Aspects (Tversky 1972) or a conjunctive or disjunctive rule. However, an alternative approach, consistent with the idea that choice processes are opportunistic, characterizes choice at a more micro level, as a sequence of operations, which are applied to the information provided to the decision maker (Huber 1980; Johnson and Payne 1985; Payne et al. 1993).

Prior research suggests two global classes of these heuristics, each characterized with a central operation: The first is screening, which employs comparisons to an external standard. Such screening is absolute, because if an alternative fails the comparison or does not have the required attribute level, it is eliminated. This operation is a central component of many well-known non-compensatory choice strategies including elimination-by-aspects, preference trees, and conjunctive and disjunctive rules. Such operations have been discussed as elimination operators or winnowing processes (Johnson and Payne 1985; McClelland et al. 1987), but they also can be used to ensure that a product has a certain characteristic. The second category of operations makes choices by comparing relative or rank-order information about attributes. Rank-order comparisons are the core mechanism behind strategies such as lexicographic procedures, majority of confirming dimensions and additive differences. At a more micro level, rank dependence has received attention as an explanatory mechanism for the departures from normative prescriptions from Value Maximization. For example, Tversky and Simonson (1993) propose a model which uses both absolute and relative evaluation to account for phenomena such as the attraction effect. Similarly, rank dependence is a major component of many descriptive theories of risky choice such as cumulative Prospect Theory and others (Tversky and Kahneman 1992).

These descriptions of choice present us with a challenge: They are applied differently across people, but also are applied at different points of a single person's choice. For example, a computer buyer whose main goal is gaming may eliminate all machines without fast graphic boards (i.e., screen), give extra weight to computers that have the fastest processor (i.e., rank order), and then trade off between brand name and price. A buyer whose main goal is word processing may have a minimum display size requirements (i.e., screen), and then tradeoff price and the kind of included software. Modeling such choice processes suggests that we must identify both who uses which operation, and when they use it.

The problem of process heterogeneity in choice models

While the notion that there is heterogeneity in processing rules both within tasks and between people may be widely-accepted among psychologists, it has been less widely embraced by analysts faced with developing formal models of choice. To understand the problem modeling rule heterogeneity poses for choice modelers consider a simple choice experiment in which a sample of consumers are asked to choose a preferred option from a number of choice sets. In each set alternatives are described by a value on each of several attributes (such as variations in a computer's price, size, and included software). In applied settings, these data are typically analyzed by assuming that the probability that alternative i will be chosen from set C by decision maker q can be described by the multinomial logit model

$$\Pr(i \mid C)_q = \frac{e^{v_{iq}}}{\sum_{j \in C} e^{v_{jq}}} \tag{3}$$

where the so-called deterministic component of utility is given by the linear-additive utility function

$$v_{iq} = \boldsymbol{\beta}'\boldsymbol{x}_{iq} \tag{4}$$

where $\boldsymbol{x}_{iq}$ is a vector of attributes of alternative i viewed by consumer q, and $\boldsymbol{\beta}$ is the associated parameter vector.

While expressions (3) and (4) offer the appeal of computational simplicity, they also form an unrealistic behavioral hypothesis about how a population of consumers would make choices. Specifically, not only are choices assumed to be made on average by maximizing a linear utility function within each choice set (something that itself might be doubted), but that this process is common to all respondents in a sample. If an analyst tries to estimate the parameters of (3) in the presence of process heterogeneity he or she thus faces a misspecification problem: the utility function required to recover the true choice probabilities in terms of (2) is of the form

$$v_{iq} = \boldsymbol{\beta}'\boldsymbol{x}_{iq} + \xi_{iCq} \tag{5}$$

where ξ_{iq} is the unobserved distortion of the strict utility of alternative i in set C caused by decision maker q using something other than a linear-compensatory utility function to make choices[1].

One approach to resolving this specification problem, of course, would be to treat ξ_{iCq} as simply another source of error in modeling, and estimate choice data using a random utility model that makes less restrictive assumptions about how the unobserved components of utility are jointly distributed across people and alternatives (such as the mixed logit or multinomial probit models; e.g., Hensher and Greene 2003). While such an approach might (albeit not certainly) allow an analyst to capture the statistical consequences of process misspecification, its obvious downside is that it does not provide a direct description of the *source* of misspecification; that is, what the rules are that are varying in a population, and how. To remedy this, several approaches to directly modeling rule variation have recently been proposed in the literature, but none have been completely satisfying in their solutions. For example, Elrod, Johnson, and White (2005), Gilbride and Allenby (2004), and Swait and Adamowicz (2001) have proposed generalized choice models that recognize the existence of a mix of compensatory and non-compensatory choice heuristics. Their limitation, however, is that they capture variation in only a small set of pre-specified heuristics (e.g., compensatory versus conjunctive screening rules) and are not easily estimated with normally-available analysis algorithms.

Process-Augmented Choice Models

In this paper we explore the viability of an alternative approach to dealing with the problem of modeling process heterogeneity. The approach allows an analyst to capture the effect of a wide range of choice rules that might arise in a task within a simple multinomial logit framework. The approach draws its heritage from the hybrid approaches to estimating conjoint models developed by Green and colleagues (e.g., Green and Krieger 1996). Like previous hybrid approaches, we seek to better inform the estimation of choice models by gathering external data on the process that individual consumers appear to be using to make choices from sets. The current approach differs, however, by using an external source of data that is gathered in the natural course of conducting a choice experiment on a computer: click-stream data measuring the frequency and sequence with which attribute information is examined by respondents. These measures form the basis of a battery of auxiliary variables designed to capture the previously-unobserved source of variance ξ_{iq} in equation (5); that is, measurement errors in revealed utility accruing to consumers using heterogeneous non-compensatory choice rules.

The approach can be more formally described as follows. Let $\boldsymbol{s}_{iq}$ be a vector of R binary (0,1) indicator variables that measure whether option i satisfies each of R individually-specified choice heuristics as defined for decision maker q. For example, one of these rules might be, "eliminate if price is more than \$800". In that case the vector element s_{irq} would take on the value 1 if option i has a price higher than \$800 (i satisfies the elimination rule) and 0 otherwise. Given a set of such measures defined for each individual, we would analyze choices using a multinomial logit model as in expression (4), but instead of the traditional linear-additive utility function defined on measured attributes as in (3), we estimate the *augmented* utility function

$$v_{iq} = \boldsymbol{\beta}'\boldsymbol{x}_{iq} + \boldsymbol{\gamma}'\boldsymbol{s}_{rq} \qquad (6)$$

where $\boldsymbol{\gamma}$ is a parameter vector that measures the incremental influence of the set of process indicators generated by the process-tracing data.

While there are a large number of heuristic policies that we could potentially attempt to measure and include in such an analysis, in this work we focus on two classes of heuristics that have been most widely identified in prior work:

1. *Screening-dependent rules (S-RULEs)*, where an option is eliminated if it either fails to possess a certain level of an attribute (a conjunctive rule) or fails to avoid a certain level (a disjunctive rule);
2. *Rank-dependent rules (R-RULEs),* where an option is either accepted or eliminated depending on whether its value on an attribute is the best or worst in a given choice set.

The vector $\boldsymbol{s}_{iq}$ thus consists of a set of binary measures of whether a given alternative satisfies the conditions of either a screening-dependent or rank-dependent heuristic.

We should emphasize that a finding of a significant effect of the process indicators in equation (5) provides a joint validation of three uncertain aspects of this analysis:

1. The ability of the process data to reveal reliable individual differences in attribute processing strategies by respondents;

2. The ability of the analyst to translate process-tracing data to indicators of specific rules; and

3. The ability these indicators to explain variance in choices beyond that provided by a standard linear-additive utility model calibrated on attribute values alone.

Like in hybrid conjoint models, equation (5) reduces to a traditional part-worth utility model if either the external measures of attribute utilization (the process indicators) are unreliable or they provide no information beyond that already provided by the linear-additive model.

Identifying Screening and Ranking Heuristics from Information-Acquisition Data

Before discussing how we construct the indicators of non-compensatory processing from process-tracing data it is necessary to first review the nature of information acquisition data as typically gathered by such procedures as manual retrieval, computer acquisition, head and eye movements. Consider again the choice experiment in which a decision maker sequentially acquires information about a set of alternatives, indexed by *i*, described by a set of attributes, here indexed by *j*. (Such an information display is illustrated in Figure 1). Over the course of the task, each cell of the information display *i,j* is acquired L_{ij} times, and each decision maker generates a series of acquisitions marked by t_{ijk}^{open} and t_{ijk}^{close}, the times at which cell *i,j* is opened and closed on the *k*th ($k=1,\ldots, L_{ij}$) acquisition. From these acquisition data we can generate a number of summary statistics (process measures) that reflect different decision heuristics (Ford et al. 1989).

In the research presented in this paper, we focus on two classes of summary statistics: those that capture the amount of *attention* received by each piece of information, and those that capture the *phase* of the decision in which an acquisition is made (for similar measures, see Ball (1997), Bockenholt and Hynan (1994), and Klayman (1985)). Here *attention* (t_{ij}) is simply defined as the total time spent looking at each cell in the information display for a given choice set:

$$t_{ij} = \sum_{l=1}^{L_{ij}} (t_{ijl}^{close} - t_{ijl}^{open}).$$

(For notational simplicity, we have suppressed s, the choice set indicator; $s = 1, \ldots, S$.) Note that for *alternatives* (pooling over attributes), low levels of $t_{i.}$ will be associated with alternatives that are eliminated in the choice process, while for *attributes* (pooling over alternatives) high levels of $t_{.j}$ will be indicative of a feature that is used for comparing alternatives (comparison).

Likewise, phase, r_{ij} is the proportion of a cell's acquisitions that occur before the timed midpoint of the decision, t^{mid}:

$$r_{ij} = \sum_{l=1}^{L_{ij}} r_{ijl} \,/\, L_{ij}$$

where $r_{ijl} = \begin{cases} 1 & if \quad t_{ijl}^{close} < t^{mid} \\ 0 & otherwise \end{cases}$

(As with the attention measure, we have suppressed the choice set indicator s for notational simplicity) For alternatives (pooling over attributes), high values of $r_{i.}$ are indicative of an alternative that was eliminated early in the choice process, while low values point to an alternative that was retained as a basis for comparison later in the process. In a similar way, for attributes, high values of $r_{.j}$ suggest that attribute j was used to eliminate or screen options, while low values suggest that the attribute was used as a basis of later comparison. These general measures of attention and phase can be used to construct the rank- and screening-dependent process indicators in equation (5).

First consider the case of screening-dependent process indicators (S-RULEs). For a given choice task involving S choice sets we wish to identify those attribute levels that are used either as a basis for a conjunctive screen ("must not have") or disjunctive screen ("must have"). There are two ways that we could make this identification, using either attention data or phase data, yielding a family of four S-RULE measures.

Let t_{l_j} be the total time spent looking at level l_j of attribute j across the S choice sets ($l_j \in \{1, \ldots, L_j\}, j = 1, \ldots, J$). We define l_j^{+a} as the feature (i.e., attribute level) that received the greatest attention of all features (i.e., $t_{l_j^{+a}} = \max(t_{l_j})$, $l_j \in \{1, \ldots, L_j\}, j = 1, \ldots, J$) and define l_j^{-a} as the feature that received the least attention (i.e., $t_{l_j^{-a}} = \min(t_{l_j})$, $l_j \in \{1, \ldots, L_j\}, j = 1, \ldots, J$).

Similarly, let r_{l_j} be the total phase for level l_j of attribute j across the S choice sets ($l_j \in \{1, \ldots, L_j\}, j = 1, \ldots, J$). We define l_j^{+p} as the feature (i.e., attribute level) with the highest total phase (i.e., $r_{l_j^{+a}} = \max(r_{l_j})$, $l_j \in \{1, \ldots, L_j\}, j = 1, \ldots, J$) and define l_j^{-p} as the feature with the lowest total phase (i.e., $r_{l_j^{-a}} = \min(r_{l_j})$, $l_j \in \{1, \ldots, L_j\}, j = 1, \ldots, J$).

We define the two "must have" or "required" S-RULE variables as

S-RULE-ReqA$_{is}$ = 1 if alternative i in choice set s includes feature l_j^{+a}, 0 otherwise.

S-RULE-ReqP$_{is}$ = 1 if alternative i in choice set s includes feature l_j^{-p}, 0 otherwise.

Similarly, we define the two "must not have" or "avoid" S-RULE variables as

S-RULE-AvoA$_{is}$ = 1 if alternative i in choice set s includes feature l_j^{-a}, 0 otherwise.

S-RULE-AvoP$_{is}$ = 1 if alternative i in choice set s includes feature l_j^{+p}, 0 otherwise.

Attention and phase data guide the construction of rank-dependent process indicators (R-RULEs) in a similar way. In this case we wish to identify whether an alternative has the maximum or minimum value in a choice set on the attribute that is attended to the most or least in the whole choice process (attention), or is looked at the latest (phase). Note that in this case we are thus defining the maxima and minima being defined at the level of *attributes* rather than attribute *levels*. If we let $t_j^{tot} = \sum_{s=1}^{S}\sum_{i=1}^{Ns} t_{ij}^{s}$ be the total time spent looking at attribute j across the S choice sets, and $r_j^{tot} = \sum_{s=1}^{S}\sum_{i=1}^{Ns} r_{ij}^{s}$ be the total phase for attribute j, we can then identify the following analogous family of R-RULE measures of a given alternative i. Let j^* be the attribute that received the most attention (i.e., $t_{j*} = \max(t_1^{tot}, t_2^{tot}, \ldots, t_J^{tot})$). We define two attention-based R-RULE variables:

R-RULE-MaxA$_{is}$ = 1 if alternative i in choice set s is best on attribute j^*, 0 otherwise.

R-RULE-MinA$_{is}$ = 1 if alternative i in choice set s is worst on attribute j^*, 0 otherwise.

Similarly, let j^+ be the attribute that tended to receive the most attention in the second half of the evaluation phase for each choice set (i.e., $r_{j+} = \min(r_1^{tot}, r_2^{tot}, ..., r_J^{tot})$). We define two phase-based R-RULE variables:

R-RULE-MaxP$_{is}$ = 1 if alternative i in choice set s best on attribute j^+, 0 otherwise.

R-RULE-MinP$_{is}$ = 1 if alternative i in choice set s is worst on attribute j^+, 0 otherwise.

Note that l_j^{+a}, l_j^{-a}, l_j^{+p}, l_j^{-p}, j^*, and j^+ are determined for each individual; as such, the values of four S-RULE and four R-RULE variables will tend to be unique to each individual, even though individuals typically face exactly the same S choice sets. With this operational definition of the R-RULE and S-RULE variables, we have a total of 8 process proxy variables.

Empirical Analysis: Can Process-Tracing Data Improve Model Performance?

Overview

We report the results of an empirical analysis exploring whether process measures enhance the descriptive and predictive performance of standard model specifications. We first describe the nature of the experimental choice data that forms the focus of the analysis, and then describe the nature of the process data that was gathered from that experiment. We then report the results of an attempt to use these data to enhance predictive validity, and compare these results to those obtained by traditional approaches to capturing individual differences in choice models. Finally, we report the results of a broader look at the way in which process data might be used to aid choice analysis in terms of guiding nested-model specifications and giving insights into the process that underlies linear-model parameters.

Data

We calibrate these models using a choice experiment examining personal computers. Personal computers were of significant relevance to many of our respondents. The subjects were fifty-nine students at an East Coast university who participated in the experiment as part of a course requirement. Each subject downloaded the software, entered a pre-assigned ID number, and was given a description of the experiment and instructions on completing the task.[2]

Respondents read instructions and practiced to familiarize themselves with the interface. They then made sixteen choices from sets of four hypothetical computer profiles, followed by two additional choices used as holdouts. Prior research and pre-testing had shown that among the most important attributes were Brand, Price, RAM and Chip (processor). Each of these attributes could take one of four values described in Table 1. The sixteen choice sets were created using a two-stage cyclical-design procedure (see, e.g., Louviere, Hensher, and Swait 2000).[3]

Results

We organize our findings into three sections. We begin by testing the basic hypothesis that we can enhance the performance of simple choice models by including process indicators (i.e., the R-RULE and S-RULES variables). Because the benchmark model for comparison in this case is rather naive—a simple homogenous logit—the emphasis is less on comparative predictive performance and more on exploring the structural properties of the proposed approach, such as the behavior of model coefficients. We then explore the viability of a latent-class and random-coefficient generalization of the proposed approach designed to capture sources of heterogeneity left unmeasured by the process measures. We close by comparing the performance of these generalizations to an alternative baseline models for non-independent choice processes, nested logit models.

The Process-Augmented Choice Model

To provide an initial view of the performance of the proposed approach, we estimated a set of conditional logit models to the choice experiment data, where the probability that a given individual would choose alternative i in from choice set C was modeled by equation (3) where the deterministic component of the utility, v_{iq}, is specified as a linear combination of the product-attributes and/or process indicators (i.e., the R-RULE and S-RULE variables). Our interest centers on the comparative performance of three alternative specifications: one defined only in terms of the product attributes, one that augments these attributes with the 8 R-RULE and S-RULE measures described above, and one defined only in terms of the 8 process indicators.

A natural concern entering into this analysis is the possibility of inter-correlations among the various process indicators, something that would arise if the heuristics used by respondents were being redundantly measured by the R-RULEs and S-RULEs. To explore this, we examined the eight process indicators to see if they provided independent information about the choice process. Inter-correlations were relatively low (.2 to .4) and a factor analysis showed no redundant structure, indicating that these four measures play complimentary roles in modeling choice processes.[4]

The results of these analyses are summarized in Table 2, which reports the derived fits and coefficients of each of the three estimated models. The data provide encouraging support for process augmentation: the basic attribute-only specification augmented with process indicators significantly improved the descriptive validity of the basic conditional logit model (a very significant improvement in fit, based on the likelihood ratio test,), with this improvement being driven by the emergence of five significant process indicators (two S-RULE and three R-RULE variables).

Perhaps just as important, the resulting model coefficients revealed strong face validity. As one would expect, the model yields a positive sign for the significant "required" measure (*S-RULE-ReqA)* and a negative sign for the significant "avoid" variable (*S-RULE-AvoA).* Likewise, we observe significant positive coefficients for terms indicating whether an option has the highest relative attribute value on the attribute that is attended to latest in the choice process (*R-RULE-MaxP*), and is attended to the most (*R-RULE-MaxA),* and negative when it has the lowest relative value on the attribute that is attended to the most *(R-RULE-MinA)*

A Broader Class of Model Comparisons

The set of process indicators could enhance predictive ability by contributing information about three potential sources of unexplained variance in choice data: individual differences in preferences that exist in a sample (e.g., preference heterogeneity), aggregate mean non-linearities in decision rules (e.g., unspecified interactions), and individual differences in these interactions (process heterogeneity, or variance in functional forms). Because the process indicators work as global aliases for these sources of variance, it is impossible to ferret from the above results the primary source of the improved fit of the process-augmented model. But we can provide at least a partial answer: if the process measures are primarily capturing preference heterogeneity, a similar level of improvement in fit would be expected to be observed from a more general, attributes-only, conditional logit that allows for preference heterogeneity, such as a latent class form (e.g., Kamakura and Russell 1989; Wedel and Kamakura 2000). Likewise, if the locus of improvement comes from capturing aggregate non-compensatory elements in the choice process, we should see a comparable improvement from a more general choice model that allows for staged processing of attributes, such as a nested logit (e.g., McFadden 1978).

We might also note that the process augmented choice model we have used is restricted, maintaining an assumption that SDUC and RDUC variables have homogeneous effects over a sample. We relax this and consider the comparative performance of a generalized version of the model that allows for at least segment-wise variation in these effects, a latent-class extension of the model given in equation (3).

Heterogeneous extensions and model comparisons

How does capturing process heterogeneity compare with capturing preference heterogeneity? To facilitate the comparison, we estimated latent-class logit models, an established approach to modeling preference heterogeneity.

In the latent-class models each individual in the sample was assumed to have a (prior) probability π_m of being a member of one of M homogeneous segments or classes, each marked by a unique coefficient vector β^m (see, e.g., Kamakura and Russell 1989). Formally, for M latent segments, the probability that a decision maker would choose alternative j from choice set C is modeled by the equation:π

$$\Pr(i \mid C)_q = \sum_{m=1}^{M} \pi_m \frac{e^{v_{iq}^m}}{\sum_{j \in C} e^{v_{jq}^m}} \tag{7}$$

where $v_{iq}{}^m = \boldsymbol{\beta}^m \mathbf{x}_{iq}$.

In addition to considering the in-sample fits of each class of models, we also examine the ability of each model to predict choice in a holdout sample. These assessments are based on two choices made by participants in the study that were not used in estimation.

We report in Table 3 the comparative model performance for the simple conditional logit model (i.e. 1-segment) and the 2- and 3-segment latent class models. The table reports the number of parameters estimated in each model as well as five fit measures: the estimation log-likelihood and associated Bayesian information criterion (BIC) number, the in-sample hit rate, and the log-likelihood and hit rate for the hold-out sample. First, the results show clear evidence that preference heterogeneity existed in the sample as implied by the increased fit of the 2- and 3-segment attributes-only latent class models over the homogeneous (1-segment) model described earlier. More importantly, however, the data also show that the simple (1-segment) process-augmented choice model describes this heterogeneity with a precision that is not dissimilar to that provided by the best attributes-only latent-class (the 2-segment solution as judged on the basis of BIC). Specifically, while the 2-segment latent-class model provides a marginally better fit in terms of pure log-likelihood (but with five extra parameters) and the difference in in-sample hit-rate is a mere 1.5% (or 1.1 percentage points).

The descriptive ability of the process-augmented model is further enhanced when we pool both approaches. Specifically, the overall best latent-class characterization of the data, as defined by in-sample log-likelihood, BIC and hit rate, is provided by a 2-segment latent-class process-augmented model. If we evaluate the alternative models on the basis of performance in the hold-out sample, the best process-augmented model — the 2-segment solution as judged on the basis of BIC — dominates the attributes-only latent-class model in terms of both log-likelihood and hit rate.

The emergence of the two-segment process-augmented model as the best apparent description of the data implies two important insights. First, it suggests that in the current sample respondents differed not only in their preferences for alternative (something captured by the purely statistical models) but also in the way attribute information was being processed—something informed by the process measures. Second, the findings also suggest that the particular way in which the process indicators affect choice is not homogeneous across the sample of respondents; while a homogeneous model provides a good basic account of the data, a generalization that allows the effects of process indicators to vary by segment offers a more complete account.

Alternative Process-augmented Models and Staged Choice Formulations

These analyses illustrate the predictive ability of only one class of augmented process models. As suggested above, process data are quite rich in structure, and it is therefore natural to ask whether other ways of constructing process indicators might provide an even better account of data. Likewise, it is also interesting to compare the performance of the process-augmented model to a random utility model that also explicitly allows for noncompensatory or staged processing of attribute information.

For the first of these comparisons, we estimated a choice model that augmented product attributes with the *raw values* of the process measures that were used to construct the process indicators above. Two such measures were considered: the total looking time directed at a given alternative, and the percentage of all acquisitions for a given option that took place in the second half of the choice process. Note that because these raw measures are observed only *ex-post,* such a model cannot be used as a forecasting tool, but rather plays the more limited role of providing a sense for the upper limit of enhanced predictive ability that process measures might be able to provide.

The results of this analysis are summarized in the first three rows of Table 4. As expected, not only do the raw process measures enhance the fit of a basic (attributes-only) choice model, but they do so beyond that indicated by the models estimated using the R-RULE and S-RULE process indicators. Specifically, a two-segment latent-class augmented model that uses the raw process measures has a BIC of 1156.5, compared to 1465.0 for the two-segment attributes plus process indicators model. Hence, while the process-indicator models do a good job capturing unobserved process heterogeneity, it understates the *potential* enhancement carried by the raw process data.

Our second interest was to assess how the performance of the process-augmented model would compare to that of a random utility model that also captured staged processing of attribute information. Because the number and breadth of such models that have been suggested in the literature is large — for example, the staged choice models by Roberts and Lattin (1991) and Swait and Ben Akiva (1987), and the cutoff models by Gilbride and Allenby (2003) and Swait (2001) — an exhaustive comparison was beyond the scope of this paper. Hence, for illustrative purposes we focus on the nested logit as the most widely-used approach to representing choice data generated by staged processes (McFadden 1978). We estimated four such models, each characterizing a two-stage choice process where the option that was worst on one of the four attributes was first probabilistically eliminated from consideration, then a choice was made among the survivors by a compensatory process. These nested models can thus be seen as implementing a homogeneous representation of a simple screening heuristic, which the process-augmented model captures through the S-RULE variables.

These analyses offer no evidence that a nested-logit model could have provided a better account of the data (middle four rows of Table 4). The fit of all four nested models to the estimation sample is inferior to that offered by any process-augmented or heterogeneous choice model. This is reflected in a mean log-likelihood for the four representations of -740 with an in-sample hit rate of 69.4%, compared to -669 and 74% for the single-segment process-augmented model described above.

The limitation of these analyses, of course, is that while there may well have been screening rules being invoked when respondents were making choices, their structures were unlikely to be homogeneous across the sample. And therein lies the argued value of the proposed process-augmented models; while one could, in theory, develop a heterogeneous system of nested representations, the process-augmented approach allows one to capture the information that would be carried by such an analysis at a far-lower computational cost.

Description and Diagnostics: Other Applications of Process Data.

We have focused on the ability of process measures to enhance the predictive performance of choice models by jointly capturing sources of preference and process heterogeneity. But improved predictive ability is not the only goal. While a random-coefficients logit calibrated just on attributes predicts about as well at the process-augmented model, we suspect it may not be as useful to a manager who wishes to explain and understand choice.

We suggest that the process data could be used in at least three additional ways to understand heterogeneity. First, they can be a useful supplemental descriptive statistic, adding to our confidence in interpreting parameters. Second, they can help diagnose areas where a linear representation may be inappropriate and evaluate possible solutions. Thirdly, they could also serve as a basis for a-priori clustering of respondents who make choices in a similar manner.

To illustrate, consider how the process data could be used to see if any particular product feature is being used in a non-compensatory fashion. If we examine the attention and phase measures, which we have used to generate R-RULEs and S-RULEs, we see that Chip is the product attribute that seems to have the most impact upon the process data, suggesting that it is likely to be used in a non-compensatory fashion.

Consistent with this is its frequent appearance in the non-compensatory process indicators. For example, we tallied how often each of the 16 attribute values was used in constructing each of the process indicators. Figure 2 represents a 3 dimensional histogram representing the frequency with which each attribute value was used as an avoid/elimination S-RULE, or as a required S-RULE (both based upon the attention measure). Another way of looking at Figure 2 is that on the left side, it tallies the number of people for whom that feature tends to be a "deal breaker" and on the right side, the number of people for whom that feature tends to be a "must-have". The figure is consistent with a number of observations. First, for the most part features are used in one role or the other. Slow chips are used to eliminate, fast chips tend to be "must-haves" (but note that the next to highest-speed chip seems to fulfill this role more often.) Second, chip is used most often as an *S-RULE*, although for some decision-makers, certain brands seem to cause elimination (Acer and E-Machines) and other are "must haves" (Compaq and Dell). Finally, there is great heterogeneity in this distribution: 8 of the 16 features serve to eliminate alternatives, and 10 of the 16 serve as must haves. This suggests that any aggregate change in functional form, such as interacting two variables, will capture only part of the non-compensatory behavior reflected in the process data.

Most importantly, process data suggests where even a state-of-the-art choice model, incorporating preference heterogeneity, might not fit the data. For example, Figure 2 suggests that because slow chips are used to eliminate alternatives, products containing those features will have predicted market shares that are biased upwards, resulting in negative residuals. Why? A compensatory model implies that other attributes, say for example, a stronger brand, will make up for a slower chip. However, if the alternative is eliminated because of a slow chip, it is, at the individual level, likely not to be chosen, no matter what the brand the computer bears. The opposite would be the case for levels used as “must-haves”, like the second fastest chip according to Figure 2.

Figure 3 and Figure 4 show the distribution of the residuals for the two-segment attributes-only and process-augmented latent-class logit models, which provided the best descriptions of the data on the basis of in- and out-of-sample BIC criteria (see Table 3). In Figure 3 there are a number of departures that indicate bias. For example, concentrating on the product attribute Chip, we see that for the slowest chip, the residuals are shifted right, reflecting a bias for over-prediction. In contrast, the higher levels of chip are left-skewed, reflecting under-prediction. Interestingly, and in addition, the variance around these predictions differs across levels: Faster chips and “better” brands seem to be characterized by much small residuals than slow chips and “weak” brands.

It is important to note that these biases have significant managerial significance when models like this are used in product design. For a manager of Acer or E-Machines, a choice model calibrated just on product attributes would over-predict their brand’s success, and suggest that a weaker brand name might be overcome by a little more RAM. A casual examination of Figure 2 suggests that for at least 20% of this sample, this conclusion is erroneous. Can the inclusion of variables that reflect screening-dependent rules (S-RULE) and rank-dependent rules (R-RULE) improve the situation? Figure 4, which shows the equivalent residual distribution for the model combining process and preference heterogeneity, suggests an improvement. There are still some reasons for concern; for example, the variance of residuals still depends upon the level of some attributes, such as processor. However, the marked skewness in Figure 3 is reduced, and the variance in residuals is also smaller, consistent with the improved fit of that model.

Extended Uses of Process-Tracing Data: Process Clustering

Finally, while the primary focus of this paper is on the ability of process indicators to capture process heterogeneity, such measures can, in principle also be used as an alternative tool for capturing preference heterogeneity. To illustrate this, one approach might be to use the process measures to define a priori decision-making segments where the parameters of a utility function are allowed to vary between groups.

We explored this idea by applying an average linkage cluster algorithm to the battery of subject-level mean process measures to derive a set of decision-process segments, and then assigned each respondent to its proximate cluster. We then estimated an attributes-only logit model for each resulting cluster. We considered 2-, 3-, and 4-segment solutions. Note that this analysis is conceptually similar to the latent-class analysis reported earlier but with two important differences: here the segments are defined *a priori* based on process measures rather than *post-hoc* using the choice data, and segment membership is deterministic, not probabilistic. The results of this analysis are reported in the bottom four rows of Table 4. The key finding is that the fit of these process-segment models is not dissimilar to that observed for the initial set of attributes-only latent class models (Table 3). Although the latent-class models provide a better fit, recall that the latent-class solution utilizes both *post-hoc* segments and probabilistic assignment—two factors that make the fit of the process-segment models encouraging.

Limitations, Extensions, and Conclusions

We have argued that heterogeneity in choice can be conceptualized as having two components: That due to individual differences in tastes (preference heterogeneity) and that due to differences in the way choices are made (process heterogeneity). We have shown that process measures, described at a finer level than in past research, can help predict choice, even in the presence of complete information about the alternatives. Finally we have explored alternative ways of incorporating information about the choice process into choice models. The best technique seems to be a latent-class model that augments traditional product-attribute measures with a set of process indicators that capture the degree to which choices appear to involve the use of rank- and screen-dependent heuristics. Such models improve in-sample fit, do a better job at predicting a holdout sample, and seem to produce less biased estimates of choice probabilities.

Caveats

When modeling individual differences in preferences, an important question concerns the ability of those estimated preferences to predict choices outside the context used for estimation. The parallel question applies to modeling preference heterogeneity: How well will it work in other contexts? While we have shown a minimal extension to a holdout sample, it is important to realize that we need to examine how well this particular technique generalizes to other decision environments. Similarly, although we think these techniques apply to many forms of process data, such demonstrations are important goals for application.

Future Research

This is obviously a first step, but one that we think provides a demonstration of the importance of modeling process heterogeneity. We demonstrate that principles based on behavioral descriptions of choice are apparently more effective than some alternative model-based solutions. However, more work is needed on two fronts: First, there needs to be more work on alternative specifications. Our notion of adding components to the current modeling technology captures some, but not all possible refinements. In particular, uniting this framework with another form of comparison, reference dependence (Hardie et al. 1993; Tversky and Kahneman 1991) would have theoretical and managerial relevance. Similarly, exploring the ability of this technology to recover coefficients and choice processes through simulation would be useful.

Likewise, one could also explore the degree to which incorporating process data might be helpful in resolving other kinds of specification problems in choice analysis that extend beyond process heterogeneity. For example, one issue that has long plagued choice models is the difficulty of separating the imbedded scale parameter that arises when estimating coefficients in multinomial logit models from the variance of the underlying utility function (e.g. Swait and Louviere 1993). To the degree that some of this variance accrues to unspecified non-compensatory processes in choice and heterogeneity in those processes, process-augmented models may prove helpful in stabilizing variances across choice environments, allowing more meaningful cross-context comparisons of model parameters than has been the case to date.

Conclusions

It has been widely argued that ignoring preference heterogeneity can lead to misleading conclusions when studying choice (Hutchinson et al. 2000), a critique that has been offered as an alternative explanation for many observed departures from rational choice theory. However, it seems that the same can be said for ignoring process heterogeneity: Not only can confirmations of rational choice succeed because they ignore process heterogeneity, but on a practical level, ignoring process can yield less accurate, and more disturbingly, biased predictions.

It appears to us that the extant literature on potential problems of compensatory models of a highly heterogeneous choice process (Andrews and Manrai 1998; Johnson et al. 1989) has had mixed impact upon the way choices are modeled: On one hand, collecting choice (as opposed to rating) data using experimental designs has become much more prevalent in both academic and commercial application. On the other hand, the primary model for analyzing such data is still a compensatory linear representation, albeit one which models individual-level preference heterogeneity through latent class or random coefficient methods. One reason for this limited impact may be the dearth of techniques that capture heterogeneity in process. This paper suggests one technique, with initial promising results. Its applicability in application is now an empirical question.

End Notes

[1]Expression (2) simply states that that for any individual choice probability $Pr(i|C)$ of the logit form there will always exist an alternative and set-specific residual ξ_{iC} that allows observed relative choice frequencies to be recovered from the linear strict utility function $v_i = \boldsymbol{\beta}'\mathbf{x}_i + \xi_{iC}$. In typical applications analysts assume $E(\xi_{iC})=0$, which yields the familiar multinomial logit (*MNL*) form. Choice processes that do not follow the *MNL* would generally require $E(\xi_{iC})\neq 0$; for example, if a decision maker uses a rank-dependent screening rule that eliminates an option if it is the lowest-ranked alternative on a particular attribute in specific choice set.

[2] Subjects read the following: In this study, you will be asked to choose which type of computer you would like to purchase. On each screen, you will see four rows and four columns. Each column describes a computer on four attributes; the attributes are listed in the rows.
Once you make a decision based on the information you have seen, click on the button along the bottom of the screen that corresponds to your choice. You will be asked to confirm your decision before moving on to the next screen. Assume that the four computers you see on each screen are identical on all but the four attributes. When you are making your choice, please choose the computer that you would be most likely to purchase from the set of four computers.

[3] The details of the procedure are as follows: First, a base set of sixteen orthogonal computer profiles were generated that reflected a main-effects fraction of a 4^4 fractional-factorial design. Sixteen choice sets were then created by taking each of these profiles and then generating three new profiles by successively folding-over attribute levels. For example, if the first alternative generated by the fractional factorial was indicating level 1 of the first attribute, level 3 of the second, level 2 of the third and level 4 of the fourth, the three other profiles in the choice set were [2,4,3,1], [3,1,4,2] and [4,2,1,3]. The order of the profiles in the choice set was randomized so it would be possible for adjacent profiles to differ by more than one increment of an attribute level

[4] The inter-correlations among process measures are sufficiently high, however, to preclude efficient estimation of model forms that recognize a complete set of interactions between process measures and product attributes. Such models would allow measurement of the degree to which tendencies to screen on certain attributes is also associated with a tendency to place greater mean weight on that attribute in compensatory trade-offs. This level of decomposition, however, does not appear easily feasible under the current approach to constructing process measures.

References

Allenby, G.M. and P.E. Rossi (1999), "Marketing models of consumer heterogeneity," Journal of Econometrics, 89 (1-2), 57-78.

Andrews, R.L. and A.K. Manrai (1998), "Simulation experiments in choice simplification experiments," Journal of Marketing Research, 35 (2), 198-209.

Ball, C. (1997), "A comparison of single-step and multiple-step transition analyses of multiattribute decision strategies," Organizational Behavior and Human Decision Processes, 69 (3), 195-204.

Bettman, J.R., E.J. Johnson, and J.W. Payne (1991), "Consumer Decision Making," in Handbook of Consumer Behavior, T.S. Robertson and H. H. Kassarjian, Eds. Englewood Cliffs, NJ: Prentice-Hall.

Bockenholt, U. and L.S. Hynan (1994), "Caveats on a Process-Tracing Measure and a Remedy," Journal of Behavioral Decision Making, 7 (2), 103-17.

Elrod, T., R.D. Johnson and J. White (2005), "A new integrated model of noncompensatory and compensatory decision strategies," Organizational Behavior and Human Decision Processes, 95 (1), 1-19.

Ford, J.K., N. Schmitt, S.L. Schechtman, and B.M. Hults (1989), "Process tracing methods: Contributions, problems, and neglected research questions," Organizational Behavior and Human Decision Processes, 43 (1), 75-117.

Gilbride, T.J. and G.M. Allenby (2004), "A choice model with conjunctive, disjunctive, and compensatory screening rules," Marketing Science, 23 (3), 391-406.

Green, P.E. and A.M. Krieger (1996), "Individualized Hybrid Models for Conjoint Analysis," Management Science, 42 (June), 850-867.

Hardie, B.G.S., E.J. Johnson, and P.S. Fader (1993), "Modeling Loss Aversion and Reference Dependence Effects on Brand Choice," Marketing Science, 12 (4), 378-94.

Hensher, David, and William H. Greene (2003), "The Mixed Logit Model: the State of Practice", *Transportation*, 30 (2), 133-176.

Huber, O. (1980), "The influence of some task variables on cognitive operations in an information-processing decision model," Acta Psychologica, 45 (1-3), 187-96.

Hutchinson, J.W., W.A. Kamakura, and J.G. Lynch (2000), "Unobserved heterogeneity as an alternative explanation for "reversal" effects in behavioral research," Journal of Consumer Research, 27 (3), 324-44.

Johnson, E.J. and R.J. Meyer (1984), “Compensatory choice models of noncompensatory processes: The effect of varying context,” Journal of Consumer Research, 11 (1), 528-41.

Johnson, E.J., R.J. Meyer, and S. Ghose (1989), “When choice models fail: Compensatory models in negatively correlated environments,” Journal of Marketing Research, 26 (3), 255-70.

Johnson, E.J. and J.W. Payne (1985), “Effort and accuracy in choice,” Management Science, 31 (4) , 395-414.

Kamakura, W.A., B.D. Kim, and J. Lee (1996), “Modeling preference and structural heterogeneity in consumer choice,” Marketing Science, 15 (2), 152-172.

Kamakura, W.A. and G. J. Russell (1989), “A probabilistic choice model for market segmentation and elasticity structure,” Journal of Marketing Research, 26 (November), 379-390.

Klayman, J. (1985), “Childrens’ decision strategies and their adaptation to task characteristics,” Organizational Behavior and Human Decision Processes, 35 (2), 179-201.

Louviere, J.J., D.A. Hensher, and Joffre Swait (2000), Stated Choice Methods: Analysis and Applications, Cambridge, U.K.: Cambridge University Press.

McClelland, G.H., B.E. Stewart, C.M. Judd, and L.E. Bourne (1987), “Effects of choice task on attribute memory,” Organizational Behavior and Human Decision Processes, 40 (2), 235-254.

McFadden, D.L. (1978), “Modelling the choice of residential location,” in Spatial Interaction Theory and Planning Models, A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull, Eds., Amsterdam: North Holland.

Payne, J.W. (1976), “Task complexity and contingent processing in decision making: An information search and protocol analysis,” Organizational Behavior and Human Decision Processes, 16, 366-387.

Payne, J.W., J.R. Bettman, and E. J. Johnson (1993), The adaptive decision maker. New York, NY: Cambridge University Press.

Revelt D. and K. Train (1998), “Mixed logit with repeated choices: Households’ choices Of appliance efficiency level,” The Review of Economics and Statistics, 80 (4), 647-657.

Roberts J.H. and J. Lattin (1991), “Development and testing of a model of consideration set formation,” Journal of Marketing Research, 28 (November), 429-440.

Swait, J. (2001), “A non-compensatory choice model incorporating attribute cutoffs,” Transportation Research B, 35 (10), 903-928.

Swait, J. and W. Adamowicz (2001), "The influence of task complexity on consumer choice: A latent class model of decision strategy switching," Journal of Consumer Research, 28 (1), 135-48.

Swait, J. and M. Ben-Akiva (1987) "Empirical test of a constrained choice discrete model: Mode choice in Sao Paulo," Transportation Research B, 21 (2), 103-115.

Swait, J. and J.J. Louviere (1993), "The role of the scale parameter in the estimation and use of multinomial logit models," Journal of Marketing Research, 30, 305-314.

Tversky, A. (1972), "Elimination by aspects: A theory of choice," Psychological Review, 79 (4), 281-299.

Tversky, A. and D. Kahneman (1991), "Loss aversion in riskless choice: A reference-dependent model," Quarterly Journal of Economics 106 (4), 1039-1061.

Tversky, A. and D. Kahneman (1992), "Advances in prospect theory: Cumulative representation of uncertainty," Journal of Risk and Uncertainty, 5 (4), 297-323.

Tversky, A. and I. Simonson (1993), "Context-dependent preferences," Management Science, 39 (10), 1179-1189.

Wedel, M. and W.A. Kamakura (2000), Market Segmentation: Conceptual and Methodological Foundations, 2nd edn., Boston, MA: Kluwer Academic Publishers.

Table 1: Study attributes and levels.[1]

Level	Brand	Price	RAM	Chip
1	Acer	$1,750	128 Meg	1.2 GHz Pentium III
2	E-Machines	$2,000	256 Meg	1.6 GHz Pentium IV
3	Compaq	$2,250	384 Meg	1.8 GHz Pentium IV
4	Dell	$2,500	512 Meg	2.0 GHz Pentium IV

[1] Values are monotonically related to values seen by respondents, but attribute values and brand names have been disguised.

Table 2: Single-Segment Conditional Logit Model Results

	Attributes-only		Process-Augmented		Process Indicators	
Variable	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value
Dell	0.18	0.16	-0.19	0.18		
Acer	-1.04	<.001	.-1.02	<.001		
Umax	-1.22	<.001	-1.39	<.001		
Price=2250	1.22	<.001	1.29	<.001		
Price=2000	1.45	<.001	1.72	<.001		
Price=1750	1.39	<.001	1.50	<.001		
RAM=384 meg	-0.60	<.001	-0.47	<.001		
RAM=256 med	-0.70	<.001	-0.57	<.001		
RAM=128 meg	-2.28	<.001	-2.23	<.001		
Chip=1.8 Ghz	-0.84	<.001	-0.66	<.001		
Chip=1.6 Ghz	-1.07	<.001	-0.94	<.001		
Chip=1.2 Ghz	-2.40	<.001	-1.91	<.001		
S-RULE-ReqA			0.62	<.001	0.30	<.001
S-RULE-ReqP			0.05	0.65	0.10	0.29
S-RULE-AvoA			-0.40	<.001	-0.36	<.001
S-RULE-AvoP			-0.16	0.18	-0.10	0.31
R-RULE-MaxA			0.44	.003	-0.75	<.001
R-RULE-MinA			-0.77	.002	-2.20	<.001
R-RULE-MaxP			0.56	<.001	0.92	<.001
R-RULE-MinP			-0.25	0.19	-1.29	<.001
LL	-736.3		-669.0		-892.1	
BIC	1555.6		1475.1		1839.1	

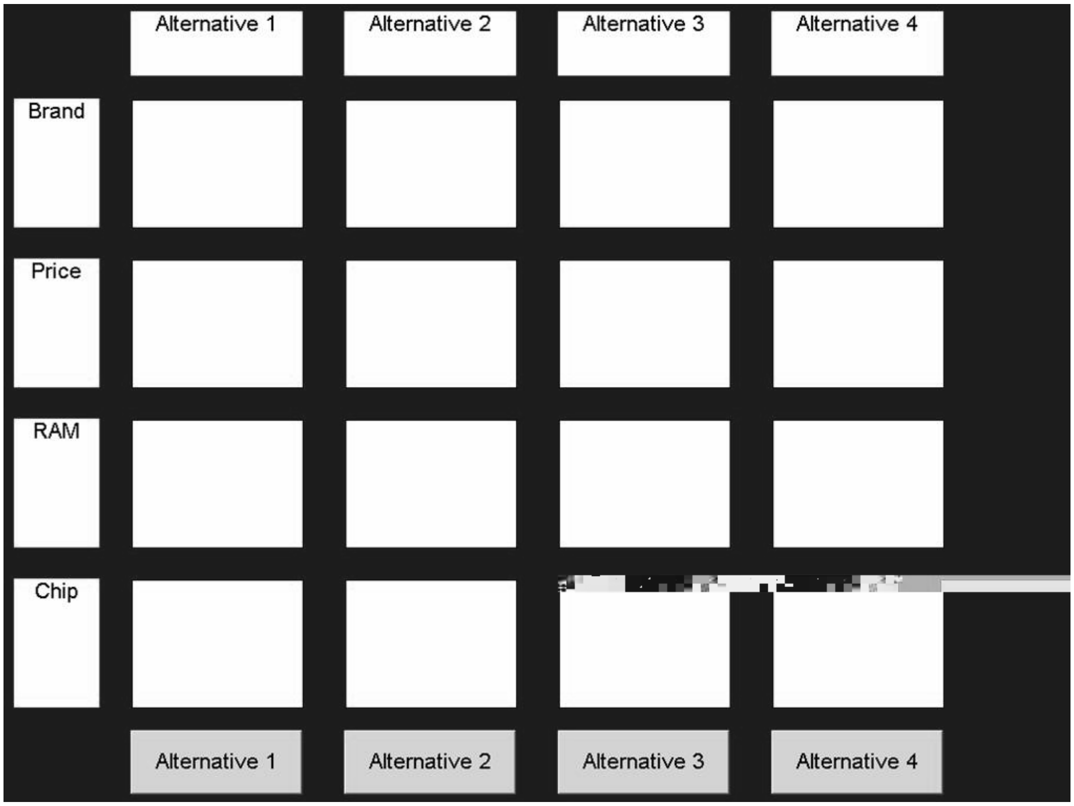
Table 3: Alternative Heterogeneous Model Fits

Model	Parameters	In-Sample LL	In-Sample BIC	In-Sample Hit Rate	Hold-Out LL	Hold-Out Hit Rate
Attributes-only Latent Class						
1-Segment	12	-736.80	1555.8	69.4%	-143.17	47.5%
2-Segment	25	-667.21	1505.7	74.9%	-138.77	51.7%
3-Segment	38	-628.97	1518.2	78.3%	-129.84	60.2%
Process-Augmented (Attributes + Process Indicators) Latent Class						
1-Segment	20	-669.04	1475.1	73.8%	-126.02	54.2%
2-Segment	41	-592.10	1465	79.8%	-114.87	61.9%
3-Segment	62	-537.34	1499.8	81.9%	-118.66	64.4%

Table 4: Raw Process Measure and Nested Logit Estimation Results

	Parameters	LL	BIC	Hit Rate
Attributes + Raw Process Measures Latent Class				
1-Segment	14	-542.9	1181.7	77.8%
2-Segment	29	-478.9	1156.5	81.3%
3-Segment	44	-437.8	1177.0	84.0%
Nested Logit				
Pruned on Brand	13	-745.6	1580.3	69.4%
Pruned on Chip	13	-743.2	1575.5	69.4%
Pruned on RAM	13	-735.3	1559.7	69.4%
Pruned on Price	13	-738.4	1565.9	69.4%
Attributes-Only Model Using Process-Derived Segments				
2-Segment	24	-716.3	1597.0	69.0%
3-Segment	36	-707.3	1661.2	69.4%
4-Segment	48	-669.1	1667.0	70.7%

Figure 1: Computer Display



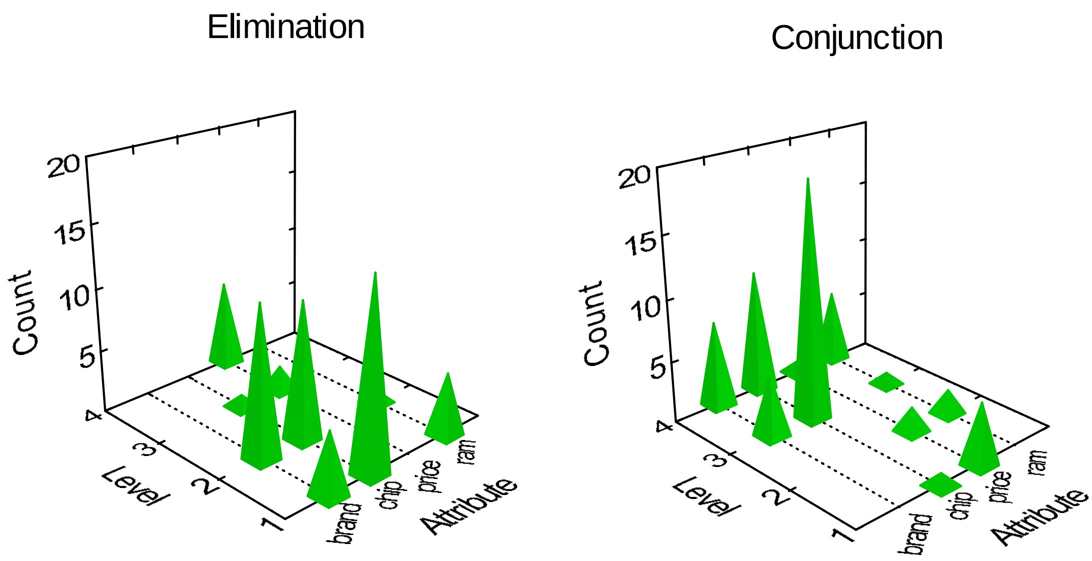


Figure 2: Distribution of Screening Dependent Utility Components (SDUCs) by Attribute and Level: Min Time (Avoid/Eliminate) and Max Time (Required)

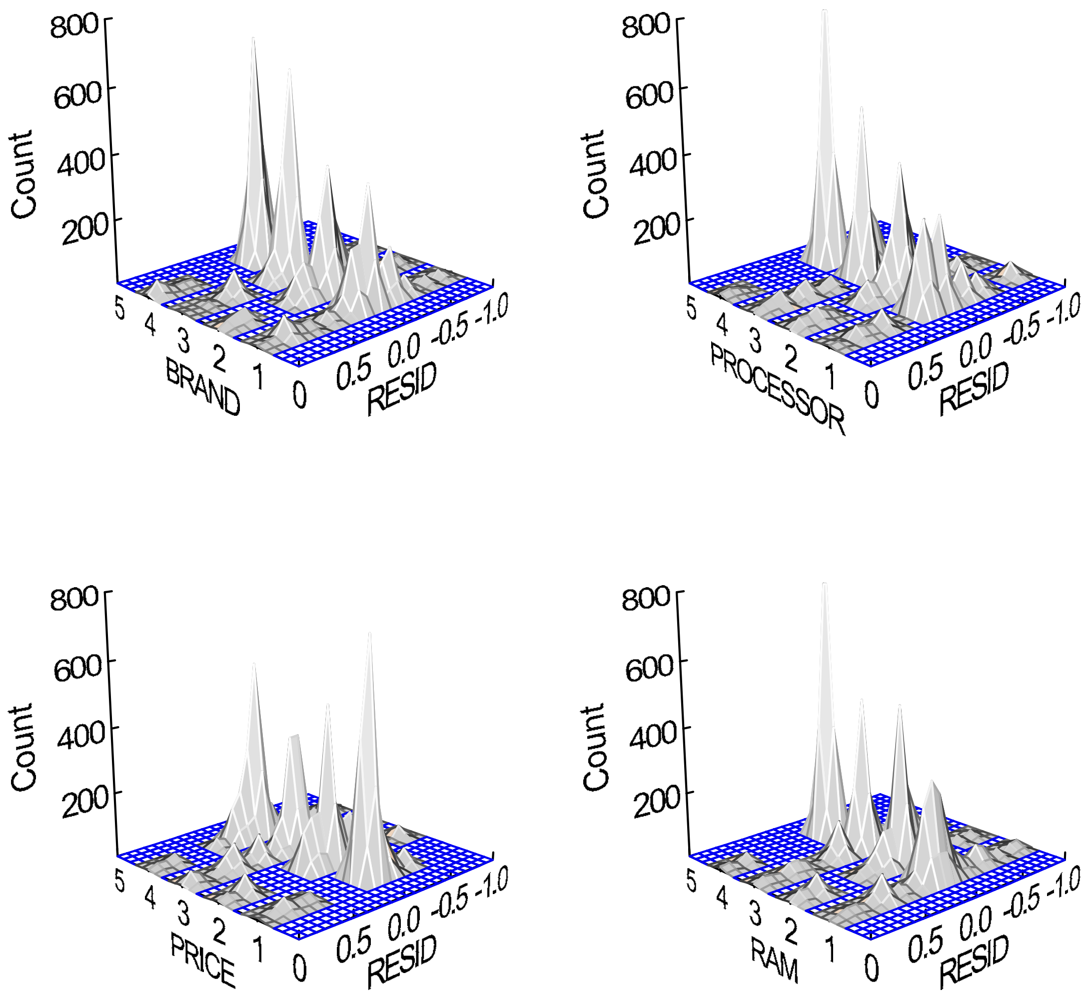


Figure 3: Distribution of Residuals by Attribute and Attribute-Level for the Two-Segment Attributes-only Latent-Class Model

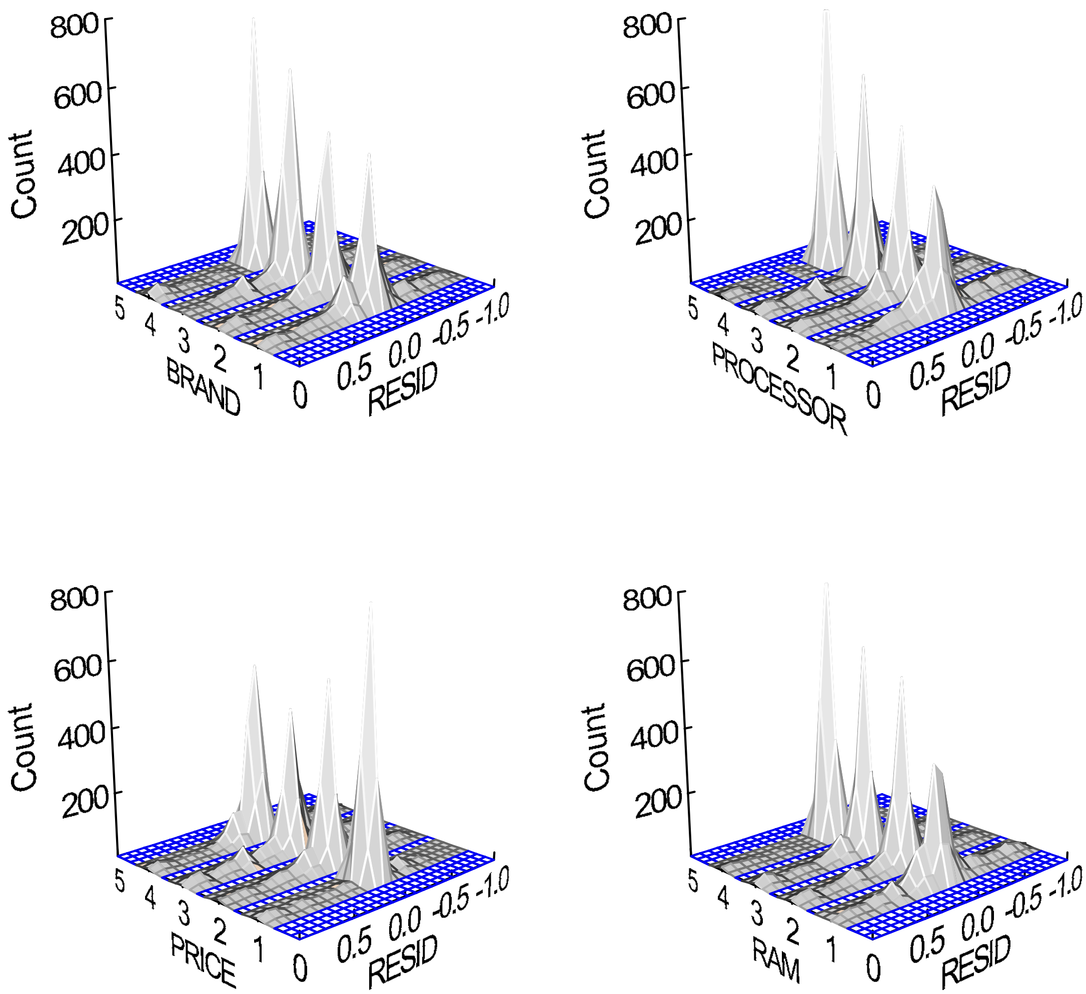


Figure 4: Distribution of Residuals by Attribute and Attribute-Level for the Two-Segment Process-Augmented Latent-Class Model