A Learning-based Model for Imputing Missing Levels in Partial Conjoint Profiles

Eric T. Bradlow, Ye Hu, Teck-Hua Ho*

March 30, 2003

*Eric T. Bradlow is an Associate Professor of Marketing and Statistics and Ye Hu is a doctoral candidate of Marketing, The Wharton School of the University of Pennsylvania. Teck-Hua Ho is William Halford, Jr. Family Professor of Marketing, Haas School of Business, University of California, Berkeley. Direct correspondence on this manuscript to Eric T. Bradlow, 3730 Walnut St., Suite 700 Jon M. Huntsman Hall, Philadelphia, PA 19104, ebradlow@wharton.upenn.edu.

Abstract

A Learning-based Model for Imputing Missing Levels in Partial Conjoint Profiles

Respondents in a conjoint experiment are sometimes presented with successive partial product profiles (i.e., profiles with missing attributes). The manner in which these respondents integrate available information in the current profile, information embedded in all previously shown profiles (perhaps through memory recall), and their prior knowledge about the product category, to impute values for missing attribute levels, is both theoretically interesting and practically relevant. Theoretically, this investigation sheds light on how customers integrate different sources of information in evaluating products with incomplete attribute information. Practically, this study highlights the potential pitfalls of imputing missing attribute levels using simple rules (e.g., an averaging model) and develops a better behavioral model for describing and predicting customers' ratings for partial conjoint profiles.

This research has two goals. First, we model how respondents infer missing levels of product attributes in a partial conjoint profile by developing a learning-based imputation model that nests several extant models. The advantage of our approach over previous research is that our general class of imputation models infers missing levels of an attribute not only from prior levels of the same attribute, but also from prior levels of other attributes (especially those that match the attribute levels of the current product profile). To account for heterogeneity in learning across individuals, we estimate this class of imputation models using a hierarchical Bayesian approach.

A second goal is to provide an empirical demonstration of our approach, and to test whether learning in conjoint studies occurs, to what extent, and in what manner it affects responses, partworths, and the relative importance of attributes. We show that the relative importance of attribute partworths can shift when subjects evaluate partial profiles. Such behavioral distortion suggests that consumers may "construct" rather than "retrieve" part-worths and hence consumers are sensitive to the way the profiles are presented. Finally, our results show that consumers' imputation process can also be influenced by manipulating their "prior" information about a product category.

Keywords: Learning Model, Pattern Matching, Multi-way Latent Contingency Table, Hierarchical Bayes

1 Introduction

Conjoint analysis is perhaps the most celebrated research tool in marketing. It has been applied to solve a wide variety of marketing problems ranging from understanding consumer preferences, estimating new product demand, to designing a brand new product line. The method involves presenting customers with a carefully chosen set of product profiles (called a test set) from the universal set (as defined by the levels of the attributes) and collecting their preferences (ratings, rankings, or choices) for product profiles in the test set. The power of the method lies in its ability to extrapolate customers' preferences from this test set to the universal set. Clearly, the conjoint analysis method works better when the test set is small and the preference task difficulty is low. Both factors can be significantly influenced by the number of product attributes.

If the number of attributes is large (as in many high-tech durable products), a full-factorial experiment would require a respondent to assess their preferences for a large number of profiles, each consisting of many attributes. The large test set problem can be solved by using a fractional design (Plackett and Burman 1946) that divides the test set among several respondents within a common customer segment.

There are two ways to solve the task difficulty problem. The first way is to use a self-explicated conjoint analysis (SECA) (Green 1984): consumers first rate the importance of the attributes, and then evaluate the attractiveness of each attribute level. By multiplying the normalized importance and attractiveness ratings, one can derive a consumer's overall preference for any profile. This approach typically requires the respondent to answer a smaller set of questions, than a full profile judgment task, and avoids the complexity of judging a profile with too many attributes. However, the SECA method has its own problems, including the fact that attribute importance ratings by respondents are not always consistent with their preference decisions; the experimental condition of separating attribute and level ratings is artificial because real-life purchase decisions are made on whole products.

The second solution is to use orthogonal subsets of all the attributes (Green 1974), the so-called partial profile conjoint analysis (PPCA). Given that profiles with a smaller number of attributes may be easier to rate[1], the PPCA approach decreases the difficulty of the rating task; however,

[1]We ran an experiment and found this hypothesis to be true for our data design; yet, it is an area for future research as to whether (and when) this is true, in general, for partial profiles.

at the same time, it may increase the number of profiles needed to get enough information about consumers' utility functions. The PPCA approach typically assumes that the attributes that are missing do not impact the product rating. Several studies cast doubt on this assumption (e.g., Feldman and Lynch 1987, Huber and McCann 1982, Broniarczyk and Alba 1994). Consequently, standard rating conjoint analysis methods, applied to partial profiles, may not produce a highly predictive utility function. This paper investigates how subjects impute missing attribate levels when evaluating partial conjoint profiles. Our goals are to understand the dependency of ratings of current profiles on all available attribute information (in both the current and previously shown profiles) including a person's prior knowledge, and to provide insights as to how consumers may impute missing levels when evaluating partial conjoint profiles.

In this paper, we relax the "null effect for missing attribute" assumption and develop a probabilistic model of how respondents impute values for missing attributes based on a person's prior for that attribute level, a given attribute's previously shown values, the previously shown values of the other attributes, and the covariation between attributes (both *a priori* and learned within the task itself). We conceptualize how consumers infer missing values via a pattern matching and learning process. Our model assumes that consumers learn after each stimulus (partial profile) about the pattern underlying the product attributes, their levels, and the correlations between them. How are strengths of patterns formed and updated? We assume that consumers have prior knowledge about the patterns and use knowledge about the product profiles acquired through the conjoint task to update their strengths. This process, like context effects in surveys (Bickart 1993, Schuman and Presser 1981), suggests that a person's partworths and the probability for an attribute level's occurrence, are not fixed and fully formed, and indeed "evolve" as the task progresses. It is this evolution (learning about the attribute level occurrences and covariation between attributes) process that we model and focus on in this paper.

We call the fundamental kernel of this updating structure a "pattern matching" learning model. That is, previously shown profiles that exhibit certain patterns among the attributes are used by the respondent to infer the missing attribute levels in the current profile. In essence, this approach can be viewed as a time-varying, multi-way contingency table of latent counts for imputation. Consequently, the *order* in which profiles are presented matters in predicting preference ratings.

1.1 An Illustrative Example

An illustrative example (Table 1) is introduced to demonstrate current extant models, the basic idea of our model and the jargon that we use. A more rigorous description follows in section 2.

Imagine a PPCA task on a product with four attributes ($A_1 - A_4$), each attribute having two levels denoted (0,1). There are three profiles ($T_1 - T_3$); each profile having one missing attribute (denoted as "?"). Following standard PPCA rating task procedures, the subjects rate T_1, T_2, T_3 sequentially. Note, for instance, that at time T_3, the respondent sees a profile with $A_1 = 1$, $A_3 = 1$, $A_4 = 1$ only. He or she does not see a "?" for A_2; Table 1 includes it only to describe the design.

[Insert Table 1 Here]

Assume that a subject has finished rating profiles T_1 and T_2, and profile T_3 is the "current" product profile. To investigate how information from different attributes might influence the value that the respondent imputes for A_2 at time T_3, we divide the attributes into three types: 1) the omni-present (OM) set; 2) the presence-manipulated (PM) set that are present (non-missing PM); and (3) the presence-manipulated set that are missing (missing PM). The OM attributes are always presented to the subjects while the PM attributes may or may not be presented. A PM attribute is called a "non-missing PM attribute" if it is *not* missing in the *current* conjoint profile. A PM attribute is called a "missing PM attribute" if it is missing in the *current* conjoint profile but may not be missing in others. In profile T_3, A_1 is an OM attribute; A_3 and A_4 are non-missing PM attributes; and A_2 is a missing PM attribute. Existing models utilize only the past information (values) from the currently missing PM attribute (A_2) for imputation of the missing level. Our model uses all three sources, missing and non-missing PM attributes and OM attributes, in which the information content is allowed to be attribute specific and to decay over time.

1.2 Link to Prior Literature

There are several different existing ways to treat missing attribute levels in the specification of utility. The first way is to assume that respondents ignore them (Green, 1974)[2]. Such an assumption implies that all the "?" in Table 1 are filled in as 0. That is, the missing level is assumed to be the level that is coded as 0 (the default level). Formally, this assumption leads to the following prediction of the missing attribute level:

$$Pr(A_2(T_3) = 0) = 1$$
$$Pr(A_2(T_3) = 1) = 0$$

Note that, in this case, the imputation process of the missing levels depends on the way the independent variables are coded (which is the dummy category), and this can be problematic theoretically.

An alternative approach is to impute a value for each of the missing attribute levels using available information. That is, we assume that people infer the levels of the missing attributes from previously shown product profiles, and weight each profile "pattern" accordingly. This latter view is consistent with Meyer (1981) where he shows that when a subject has no information about certain attributes, that attribute is not ignored, but rather is assigned a score equal to the individual's adaptation level.

There are two common ways to model how consumers make inferences about missing attribute levels. One way is based on the so-called "recency effect" (Lynch and Srull 1982): people assume the missing attribute level to be the last level of the same attribute they saw. According to such a model, in Table 1, A_2 in profile T_3 takes level 1 (following the level of A_2 in profile T_2). Formally, we have:

$$Pr(A_2(T_3) = 0) = 0$$
$$Pr(A_2(T_3) = 1) = 1$$

A second commonly used imputation approach is the averaging model (Yamagishi and Hill 1981). In this model, people impute the missing attribute level by averaging all the previously shown

[2]Note that the assumption of ignoring the missing attributes in pair-wise conjoint partial profiles corresponds to the cancellation of attributes not shown.

levels of the missing attribute. For example, in Table 1, we have[3]:

$$Pr(A_2(T_3) = 0) = \frac{1}{2}$$
$$Pr(A_2(T_3) = 1) = \frac{1}{2}$$

In imputing the missing values, the recency and averaging models make strong assumptions about the similarity between the current profile and the previous profiles. The recency-based model assumes that the current profile is "similar" only to the most recently shown profile and is dissimilar to the rest[4]. The averaging model assumes that the current profile is equally similar to all the previously shown profiles. One would expect that some previously shown profiles are more similar ("count more" in imputing) to the current profile while others are less so.

A simple way to relax the strong assumptions made in the recency and averaging models is to weight more recent information more heavily than less recent information by discounting or decaying past information by a factor per each elapsed time period. Additionally, and as described fully in the next section, in all the above models, only the information from the missing PM attribute itself is used to impute the level of the missing PM attribute. As we demonstrate empirically, and via covariation one could imagine conceptually, that correlation patterns among the attributes can influence a given attribute's imputed value. Hence, in many ways, the averaging and recency models may be an improvement over ignoring the missing attribute levels, but still are quite deficient. It is an extension of these basic models that we consider here.

1.3 Using Covariation Among Attributes

All the above models utilize *only* past data from A_2 to impute A_2 in profile T_3. By doing so, they ignore two important pieces of information. First, there are the complete set of patterns defined by previous product profiles shown to the subjects (T_1, T_2), not just the values for A_2 ($A_2(T_1)$, $A_2(T_2)$) (i.e. utilize the missing PM attribute only). Some of these patterns might occur more frequently than others, so their attribute level values for A_2 might be more salient and memorable.

[3]Note that technically the averaging model is not stochastic and would have $A_2(T_3) = \frac{1}{2}$. We implement its stochastic version here.

[4]Note that if T_2 were to be first profile and T_1 were to be the second profile, the prediction of "recency model" would have been reversed (i.e., $Prob(A_2 = 0|T_3) = 1$). The averaging model, however, would have given the same prediction.

Second, there is the current partial profile (profile T_3). For example, using a digital camera, as in section 4, if high price is highly correlated with a given camera size, then utilizing the covariation information across attributes provides valuable information. Such a correlation structure could be based on people's long-term memory or from the "learning" in the conjoint task. Huber and McCann (1982) showed that people use their belief of the correlation structure between price and quality to infer the missing price or quality when either one is missing. Broniarczyk and Alba (1994) also show that consumers' intuitions (priors) influence their inference making. Our model captures these covariances as well as the priors they "arrive" at the experiment with, in a parsimonious way.

The rest of this paper is composed of four sections. In section 2, we develop our basic model. In section 3, the design of our experiment is explained. In section 4, we describe an empirical application of our model. Section 5 contains conclusions, caveats, and future research.

2 The Imputation Model

2.1 Basic Ideas

First, we use Table 1 and attribute 2 (which is missing) at time 3, to demonstrate the basic idea of our model. What we are trying to infer from the imputation model is the probability that the missing PM attribute A_2 in profile T_3 takes value 1 or 0. The parameterization of the probabilities is based upon the work of Camerer and Ho (1999), which describes how information decay (in our case pattern reinforcement) occurs over time, and Hoch, Bradlow and Wansink (1999), which describes an approach to measure the similarity between pairs of profiles. The two basic concepts we provide are what we call "Pattern Matching" and "Experience Counts". They are then combined to form our imputation model which defines the probabilities over the missing attribute levels.

Three potential classes of models (Figures 1A-1C) are provided to demonstrate how the generality of our model is built up sequentially, using varying information sources.

[Insert Figures 1A, 1B, and 1C Here.]

The model in Figure 1A uses only previously shown information about the missing PM attribute (i.e., A_2) to impute missing levels. It is a natural extension of the recency and averaging model but allowing for decay. The imputation is based on historical levels of A_2 (0 in profile T_1 and 1 in profile T_2), but now with the more recent level (1 in profile T_2) being potentially more influential. To capture the "recency effect" in a "decay-weighed" averaging model, we introduce a *decay* parameter, λ_2 ($0 < \lambda_2 \leq 1$, the subscript denoting it is for A_2). We also introduce a new concept, "experience count (EC)" (Camerer and Ho, 1999) such that $N_2(3|0)$ denotes the EC of A_2 in the 3^{rd} profile, taking level 0.

As illustrated in Figure 1A, $N_2(3|0)$ "adds up" (is only influenced by) level 0 of A_2 in profile T_1 as if $A_2(T_3)$=0, it matches only $A_2(T_1)$. Thus we have $\mathrm{N}_2(3|0) = \lambda_2^2$. λ_2 takes a power of 2 because there are two periods of difference between profile T_1 and profile T_3. Simply put, attribute information used for imputation has a geometric decay. Similarly, we have $N_2(3|1) = \lambda_2$; profile 2 has $A_2(T_2) = 1$. The probability that $A_2(T_3)$ is imputed to take levels 1 or 0 are thus:

$$\begin{aligned}
\Pr(A_2(T_3) = 1) &= \frac{N_2(3|1)}{N_2(3|1) + N_2(3|0)} = \frac{\lambda_2}{\lambda_2 + \lambda_2^2} \\
\Pr(A_2(T_3) = 0) &= 1 - \Pr(A_2(T_3) = 1)
\end{aligned}$$

where $A_2(T_3)$ denotes the level of A_2 in profile T_3. The averaging model is a special case of this class of models in which $\lambda_2 = 1$. It predicts that $\Pr(A_2(T_3) = 0) = \Pr(A_2(T_3) = 1) = 0.5$. The recency model is captured by (nested by our model) having $\lambda_2 \to 0$, with corresponding probabilities $\Pr(A_2(T_3) = 1) \to 1$ and $\Pr(A_2(T_3) = 0) \to 0$.

The models in Figures 1B and 1C, extensions of the model in Figure 1A, are based upon what we term conditonal "pattern matching". In our model, a "pattern" is based on a pair of attributes, with one of them being a missing PM attribute, the other being either a non-missing PM or OM attribute. Two patterns conditionally match up when the levels of both pairs of attributes match between profiles.

The model in Figure 1B uses information from both the missing and non-missing PM attributes to impute the missing PM attribute level. That is, in addition to using information from A_2 itself, as in the model in Figure 1A, we utilize possible conditional match patterns between profiles of the non-missing PM attributes (A_3, A_4)[5]. We assume that the missing attribute levels ($A_3(T_1)$ and $A_4(T_2)$) are not used for imputation. Thus, we need only to check whether there is a match

[5]Note, we require that the levels of the previous profiles match with the non-missing PM levels of the current profile. For simplicity we assume that previously imputed levels are not used for imputation. This is a limitation to consider for future research.

between $[A_3(T_3)]$ and $[A_3(T_2)]$ and between $[A_4(T_3)]$ and $[A_4(T_1)]$. Since there is a match between $[A_3(T_3)]$ and $[A_3(T_2)]$ (both equal to 1), we expect T_2 to influence imputation more and hence the probability that $A_2(T_3)$ is imputed as the same level as $A_2(T_2)$ (i.e., 1) increases. Another decay parameter, λ_3 ($0 < \lambda_3 \leq 1$), is added to capture this reinforcement from non-missing PM attribute A_3 on the imputation of A_2. Consequently, with the additional term λ_3^1 (with power of 1 because of 1 period of difference between profiles T_3 and T_2), $N_2(3|1)$ becomes $\lambda_2 + \lambda_3$. $N_2(3|0)$ stays the same, still equaling λ_2^2, as in Figure 1A. The probabilities that $A_2(T_3)$ is imputed as 0 or 1 now becomes:

$$\begin{aligned}
\Pr(A_2(T_3) = 1) &= \frac{N_2(3|1)}{N_2(3|1) + N_2(3|0)} = \frac{\lambda_2 + \lambda_3}{(\lambda_2 + \lambda_3) + \lambda_2^2} \\
\Pr(A_2(T_3) = 0) &= 1 - \Pr(A_2(T_3) = 1)
\end{aligned}$$

Notice that $Pr(A_2(T_3) = 1)$ becomes larger, when compared to the previous class of models. This makes sense because T_2 (when compared to T_1) is more similar to T_3 when we consider only PM attributes.

The model in Figure 1C uses all available information of both the PM (missing as well as non-missing) and OM attributes to impute the missing level. Following the procedure above, the ECs are $N_2(3|0) = \lambda_1^2 + \lambda_2^2$ and $N_2(3|1) = \lambda_2 + \lambda_3$, where λ_1^2 indicates that there is a match between $[A_1(T_3)]$ and $[A_1(T_1)]$ (both equal to 1). The corresponding probabilities now become:

$$\begin{aligned}
\Pr(A_2(T_3) = 1) &= \frac{N_2(3|1)}{N_2(3|1) + N_2(3|0)} = \frac{\lambda_2 + \lambda_3}{(\lambda_1^2 + \lambda_2^2) + (\lambda_2 + \lambda_3)} \\
\Pr(A_2(T_3) = 0) &= 1 - \Pr(A_2(T_3) = 1)
\end{aligned}$$

Obviously, now T_3 looks less like T_2 because of a match between T_3 and T_1 on attribute 1. So, $\Pr(A_2(T_3) = 1)$ becomes smaller compared to the case in Figure 1B. In the next section, we formally parameterize the imputation model.

The model presented in Figure 1C has two desirable properties:

1. It uses all available information in the previously shown and current profiles in a sensible way. The model highlights the potential pitfalls of the averaging and recency models. For example, it implies that these previous simpler models will yield the same prediction, given above, if T_3 were to take any of these patterns $\{[1, ?, 1, 1], [1, ?, 0, 0], [1, ?, 0, 1], [0, ?, 1, 0], [0, ?, 0, 0], [1, ?, 1, 0], [1, ?, 1, 0]\}$, which seems very unlikely.

2. It allows respondents to apply different weights to different attributes[6] depending on their preference. For example, a respondent who has a good knowledge of, or puts a strong emphasis on, price may use it more heavily than other attributes in imputing the missing attribute level.

2.2 General Formulation

In a rating-task conjoint experiment, we investigate how I individuals (indexed by $i = 1, \cdots, I$) rate a series of product profiles (partial or full). Each product profile is characterized by J attributes (indexed by $j = 1, \cdots, J$) and each attribute j has K_j levels (indexed by $k_j = 1, \cdots, K_j$)[7]. Each individual i rates T profiles (indexed by $t = 1, \cdots, T$) one by one. Individual i's rating for product profile t is denoted by $y_i(t)$.

We denote the stimulus and design matrices faced by each subject i in a conjoint experiment as $\boldsymbol{X}_i$ and $\boldsymbol{R}_i$. $\boldsymbol{X}_i$ is the matrix of product attribute levels, with a dimension of $T \times \sum_{j=1}^{J}(K_j - 1)$ and elements $x_{ijk_j}(t)$ $(j = 1, ..., J, k_j = 1, ..., K_j - 1)$. $\boldsymbol{R}_i$ has a dimension of $T \times J$ with elements $r_{ij}(t)$. An attribute is denoted as a row vector $\boldsymbol{x}_{ij}(t) = \left[x_{ij1}(t), x_{ij2}(t), \cdots, x_{ij(K_j-1)}(t)\right]$. Note that when attribute j is missing, it is the entire vector $\boldsymbol{x}_{ij}(t)$, that is missing. Each product profile is denoted as a row vector $\boldsymbol{x}_i(t) = [\boldsymbol{x}_{i1}(t), \boldsymbol{x}_{i2}(t), \cdots, \boldsymbol{x}_{iJ}(t)]$. The design matrix $\boldsymbol{R}_i$ indicates which attribute(s) are missing. For example, $r_{ij}(t) = 1$ when $\boldsymbol{x}_{ij}(t)$ is present; $r_{ij}(t) = 0$ otherwise.

From the imputation model, we construct an imputed stimulus matrix $\boldsymbol{X}'_i$ (the "prime" is used as a superscript, not matrix transposition or derivatives) from the stimulus matrix $\boldsymbol{X}_i$ where the elements of the former are derived from those of the latter as follows:

$$\boldsymbol{x}'_{ij}(t) = \begin{cases} \boldsymbol{x}_{ij}(t) & \text{if } r_{ij}(t) = 1 \\ \boldsymbol{I}\boldsymbol{x}_{ij}(t) & \text{if } r_{ij}(t) = 0 \end{cases}$$

The imputation modeling effort is to define the sequence of time ordered probabilities for the missing attribute levels, i.e., to determine $\Pr(\boldsymbol{x}'_{ij}(t))$ when $r_{ij}(t) = 0$. These probabilities can

[6] As given in equation (1) and Section 4, we allow the λ to be individual and attribute specific.

[7] One assumption underlying our approach is that respondents "know" the universe of attributes and their corresponding levels. This assumption may not be very restrictive in our task in which we have a learning phase. Note that if this does not hold in practice, then we simply restrict our model to those levels and attributes they know. If attribute j is a continuous variable, it is treated as having K_j discrete levels with $K_j - 1$ dummy variables. However, a distance measure such as Euclidean distance, instead of attribute level match-up, should be used in pattern matching.

then be combined with the complete data likelihood to yield a mixture model over the feasible states (possible profiles).

2.2.1 Pattern Matching

To construct the probability $\Pr(\boldsymbol{x}'_{ij}(t))$, we first parameterize and describe the pattern matching approach from our examples above. The match-up of a given pair of attributes [Yes/No] between two profiles is defined as follows. Assume $r_{ij}(t) = 0$ and the level of attribute j is to be imputed. We match up (assess Yes/No, do they match?) the patterns of $[\boldsymbol{x}'_{ij}(t), \boldsymbol{x}_{ij'}(t)]$ and $[\boldsymbol{x}_{ij}(t'), \boldsymbol{x}_{ij'}(t')]$, where $j' = 1, ..., J, j' \neq j, t' = 1, ..., t-1$. That is, the *imputed* level of attribute j in current profile t has to match the level of attribute j in profile $t' < t$, i.e., $\boldsymbol{x}'_{ij}(t) = \boldsymbol{x}_{ij}(t')$ for that match to count. We call this a conditional match-up model, as the pairs of profiles must match on the missing PM attribute. Note, that we fit our entire sequence of models with and without this conditional restriction, yielding substantively identical results.

Conditional on matching the missing PM, i.e. our conditional match model, two patterns match when $\boldsymbol{x}_{ij'}(t) = \boldsymbol{x}_{ij'}(t')$. These two steps are formulated as $1\left[\boldsymbol{x}_{ij'}(t) = \boldsymbol{x}_{ij'}(t') \middle| \boldsymbol{x}'_{ij}(t) = \boldsymbol{x}_{ij}(t')\right]$, where $1(\cdot)$ is an indicator function.

It is also important to note the following "properties" of our pattern matching approach. (1) We do *not* match profiles based on imputed values of previous attributes, or in a more-than-one missing attribute case, imputed values of missing PM attribute(s) $j' \neq j$. (2) The way in which we "match" a given pattern is binary (yes it matched, no it did not). One could imagine a metric-based degree of matching model. We chose a binary Hamming metric approach as it is parsimonious, easy to describe, and is cognitively simple. However, in the conclusion section we suggest that a model which combines conjoint analysis and Multidimensional Scaling (MDS) methodologies is an important future research area.

2.2.2 Experience Counts

As shown with the examples in Figures 1A-1C, there are two basic tenets of our approach which clearly extends common extant models for imputation of missing attribute levels. First, matched patterns from OM attributes and non-missing PM attributes have the potential to reinforce the pattern in the *current* profile. Secondly, that the effects of pattern reinforcement are "time dependent" (i.e. can decay). That is, a pattern which matched on a previous profile may have greater reinforcement weight than patterns which match based on profiles further in the past. Such an assumption is common in learning (Camerer and Ho 1999) and is applied here. This allows us to capture memory decay and the tendency of subjects to remember more recent patterns.

Similar to our example, let $N_{ijk_j}(t|\boldsymbol{x}'_{ij}(t) = \boldsymbol{l}_{jk_j})$ ($k_j = 1, ..., K_j$, where K_j represents the base level of dummy coded attribute j) denote the latent "experience count" of person i for attribute j to take level k_j at time t. Further, let $\boldsymbol{l}_{jk_j}$ denote a row vector of size $K_j - 1$, with the k_j^{th} element taking value 1 and the other elements taking value 0 when $k_j = 1, ..., K_j - 1$, and all elements taking value 0 when $k_j = K_j$. With attribute j as the missing PM attribute, our model for $N_{ijk_j}(t|\boldsymbol{x}'_{ij}(t) = \boldsymbol{l}_{jk_j})$ is given by

$$N_{ijk_j}(t|\boldsymbol{x}'_{ij}(t) = \boldsymbol{l}_{jk_j}) = N_{ijk_j}(0|\boldsymbol{l}_{jk_j}) + \qquad (1)$$

$$\sum_{t'=1}^{t-1} \left\{ \lambda_{i(j,j)}^{t-t'} \cdot 1[\boldsymbol{x}'_{ij}(t) = \boldsymbol{x}_{ij}(t')] + \sum_{j'=1, j' \neq j}^{J} \left[\lambda_{i(j,j')}^{t-t'} \cdot 1 \left[\boldsymbol{x}_{ij'}(t) = \boldsymbol{x}_{ij'}(t') \left| \boldsymbol{x}'_{ij}(t) = \boldsymbol{x}_{ij}(t') \right. \right] \right] \right\}$$

= "prior count" + {"missing PM count" + ["non-missing PM attribute + OM attribute count"]} where $N_{ijk_j}(0|\boldsymbol{l}_{jk_j})$ denotes the "prior" count of person i on attribute j, level k_j at "time 0"; $0 < \lambda_{i(j,j')} \leq 1$ is the decay parameter of person i relating attribute j' ($j' = 1, ..., J$) to attribute j. $N_{ijk_j}(0|\boldsymbol{l}_{jk_j})$ allows for the possibility of prior knowledge of the marginal frequency of attribute levels and prior correlation between attribute levels.

In our experimental results, section 4, we fit a fairly general model that has the following set of specifications for $\lambda_{i(j,j')}$. This corresponds to Table 3, an example with digital cameras in which $j = \{1, 2, 3, 4\}$ are PM attributes (delay between shots, storage media, maximum resolution, and camera size) and $j = \{5, 6\}$ are OM attributes (price and mini-movie).

$$\begin{aligned}
\lambda_{i(j,j')} &= \lambda_{ij} \quad \text{if } j' = j, j' \in PM; \ \forall j = 1, 2, 3, 4 \\
\lambda_{i(j,j')} &= \lambda_{i5} \quad \text{if } j' \neq j, j' \in PM \\
\lambda_{i(j,j')} &= \lambda_{i6} \quad \text{if } j = \text{Maximum Resolution and } j' = \text{Price}
\end{aligned}$$

Note that we assume that λ_{im} ($m = 1, ..., 6$) does not vary with t, i.e., we assume a constant rate of decay over the entire profile set. Certainly a future area of research would be to allow for a varying decay rate. That is, there is empirical evidence in extant studies (e.g. West, Brown, and Hoch 1996) to suggest that people's attribute learning, as in our model, may occur for say the first T^* profiles, but then stabilize after that. λ_{i6} is included in the model, as described in section 4, due to a prior manipulation of the covariance between price and maximum resolution.

We describe in the next subsection how the values of $N_{ijk_j}(t|\boldsymbol{x}'_{ij}(t) = \boldsymbol{l}_{jk_j})$ are used to compute $Pr(\boldsymbol{x}'_{ij}(t))$, the probability distribution over the missing attribute levels.

2.2.3 Probability

The structure described in section 2.2.2 defines the entire imputation process for partial profile conjoint designs as a time varying latent contingency table with counts $N_{ijk_j}(t|\boldsymbol{x}'_{ij}(t) = \boldsymbol{l}_{jk_j})$ given in Equation (1). Hence, following our example, $\Pr(\boldsymbol{x}'_{ij}(t) = \boldsymbol{l}_{jk_j})$ is given by

$$\Pr\left(\boldsymbol{x}'_{ij}(t) = \boldsymbol{l}_{jk_j}\right) = \begin{cases} 1 & \text{if } r_{ij}(t) = 1 \text{ and } \boldsymbol{x}'_{ij}(t) = \boldsymbol{l}_{jk_j} \\ 0 & \text{if } r_{ij}(t) = 1 \text{ and } \boldsymbol{x}'_{ij}(t) \neq \boldsymbol{l}_{jk_j} \\ \frac{N_{ijk_j}(t|\boldsymbol{x}'_{ij}(t)=\boldsymbol{l}_{jk_j})}{\sum_{k'_j=1}^{K_j} N_{ijk'_j}(t|\boldsymbol{x}'_{ij}(t)=\boldsymbol{l}_{jk'_j})} & \text{if } r_{ij}(t) = 0 \end{cases} \tag{2}$$

That is, the probability a given attribute level is imputed, when that attribute is missing, is its "proportion" of the total experience count for its attribute levels. As $N_{ijk_j}(t|\boldsymbol{x}'_{ij}(t) = \boldsymbol{l}_{jk_j})$ incorporates information across patterns to reinforce each pattern, and allows for differing importance across time, this model satisfies our basic pattern matching and reinforcement requirements. When there are multiple attributes missing, we assume independence of counts to derive the joint probability of a pattern missing; however the counts are correlated as attribute levels that occur together have counts that will be updated together, and prior counts that are related.

2.3 Testing for Imputation in a Rating Conjoint Setup

In this section, we integrate our imputation model into the formulation of a standard conjoint analysis. There are four sources of information that can be used for inference making about missing attribute levels.

(1) Initial prior on attribute levels, $N_{ijk_j}(0|\boldsymbol{l}_{jk_j})$

(2) "Cross-time information", or, previously shown profiles $(\boldsymbol{x}_i(1), \cdots, \boldsymbol{x}_i(t-1))$.

(3) The current partial profile, $\boldsymbol{x}_i(t)$.

(4) The previous ratings $y_i(1), \ldots, y_i(t-1)$.

Our model for $\Pr(\boldsymbol{x}'_i(t))$ will incorporate the first three of these information sources. To control for the possible effect that individuals may anchor on previous ratings that they have given (Tversky and Kahneman 1974; Wilson et. al. 1996), we use a first-order autoregressive process (AR(1)) for the lagged residuals (see Equation (3)).

Another facet that we have incorporated into the model is based on the works of Johnson and others (Johnson 1987, Johnson and Levin 1985, Levin et al 1986). Their work suggests that subjects have different partworths for the same attribute when it's missing compared to when it is not. Such differences, at the same time, may interact with the difference in product attribute levels.

In Table 2, we set two partworths for each attribute(in the case of K_j=2; our example in section 4), which allows us to statistically test the hypotheses of whether the β's are significantly different from zero. β_{ijk_j} corresponds to the case where $x'_{ijk_j}(t) = 1$ (regardless of whether it was observed or imputed); β'_{ijk_j} refers to whether the partworth for an attribute-level increases or decreases when that attribute is shown (i.e., $r_{ij}(t) = 1$) or not[8]. Note, that without an imputation process, one could not disentangle the partworths into their impact for level and missingness since the full 2 x 2 table, without imputation, is not observed. We believe that this integration of the work of Johnson and others, directly into our imputation model, represents a further significant contribution of this work.

[Insert Table 2 Here.]

The null hypotheses of interest, that we test in section 4 are $\beta_{ijk_j} = 0$ and $\beta'_{ijk_j} = 0$.

Now the conjoint model becomes:

$$y_i(t) \quad = \quad \alpha_i + \sum_{j=1}^{J} \sum_{k_j=1}^{K_j-1} \left[\beta_{ijk_j} x'_{ijk_j}(t) + \beta'_{ijk_j} r_{ij}(t) \right] + \gamma_i \epsilon_{i(t-1)} + \epsilon_{it} \tag{3}$$

where we assume that $\epsilon_{i0} = 0, \forall i$. The probabilities $\Pr(\boldsymbol{x}'_{ij}(t) = \boldsymbol{l}_{jk_j})$ from equation (2) provide information on the values taken by the $x'_{ijk_j}(t)$'s in equation (3).

To both accommodate subject heterogeneity, we utilize a hierarchical Bayesian (HB) approach to estimate the model. The setup of HB estimation is explained in the next section.

[8]We've run models with the interaction of the two — when the attribute is shown and $x'_{ijk_j}(t) = 1$, and found that the interaction partworths are *not* significant. Thus, those interaction partworths are not included in our model. However, we believe that in future empirical studies understanding the dynamics of partworth changes as a function of their level, when imputed, will be critical.

2.4 Distribution of Heterogeneity

We allow for the rate of information decay for a specific attribute pattern to be individual and pattern (attribute) specific recognizing that considerable heterogeneity is likely to exist across persons in their decay attribute imputation parameters, λ_{im} $(m = 1, ...6)$. In addition, the basic parameters of the conjoint model, the individual conjoint intercepts, α_i, the attribute partworths, β_{ijk_j} and β'_{ijk_j}, and the residual variances, $var(\epsilon_{it}) = \sigma_i^2$, may also contain considerable heterogeneity, yet share commonalities across the population of inference. To account for this heterogeneity in a coherent fashion, we nest our model in a Bayesian framework (Gelfand and Smith, 1990). Denoting:

$$\Delta_i(t) = y_i(t) - \left\{ \alpha_i + \sum_{j=1}^{J} \sum_{k_j=1}^{K_j-1} \left[\beta_{ijk_j} x'_{ijk_j}(t) + \beta'_{ijk_j} r_{ij}(t) \right] \right\}$$

our complete model specification is given by[9] (assume that $\Delta_i(0) = 0$):

$$\Delta_i(t) - \gamma_i \Delta_i(t-1) \sim N\left(0, \sigma_i^2\right) \tag{4}$$

with prior specifications for the conjoint parameters given by $(\forall i, j, k_j)$

$$\begin{aligned}
\gamma_i &\sim U(-1, 1) \\
\alpha_i &\sim N(\bar{\alpha}, \sigma_\alpha^2) \\
\beta_{ijk_j} &\sim N(\bar{\beta}_{jk_j}, \sigma^2_{jk_j\beta}) \\
\beta'_{ijk_j} &\sim N(\bar{\beta}'_{jk_j}, \sigma^2_{jk_j\beta'}) \\
\sigma^2_\alpha, \sigma^2_{jk_j\beta}, \sigma^2_{jk_j\beta'} &\sim \text{Inv-}\Gamma(\cdot, \cdot),
\end{aligned}$$

and prior specification for the attribute decay parameters, $0 < \lambda_{im} \leq 1$, given by

$$\lambda_{im} \sim Beta(a_m, b_m). \tag{5}$$

We note that a more general result would allow for a multivariate prior; however, we restrict here to independence to reduce the number of parameters. The prior ECs of each individual are assumed to follow a Poisson distribution with parameter varying by individuals and attribute levels:

$$N_{ijk_j}(0|\boldsymbol{l}_{jk_j}) \sim Poisson(\exp(\zeta_i + \omega_{jk_j})), \tag{6}$$

[9] We would like to thank one of the anonymous reviewers for suggesting this specfication.

with slightly informative priors on ζ_i and ω_{jk_j}. We note that $N(x, y)$ denotes a normal distribution with mean vector x and variance y, $U(x, y)$ a uniform distribution with lower bound x and upper bound y, Inv-$\Gamma(\cdot, \cdot)$ an inverse gamma distribution with corresponding parameters and $Beta(a, b)$ a beta distribution with parameters a, b. To complete the model specification, slightly informative hyperpriors were placed on σ^2_α, $\sigma^2_{jk_j\beta}$, and $\sigma^2_{jk_j\beta'}$, $\forall j, k_j$. (an inverse-gamma distribution with 0.0001 shape and scale), $\bar{\beta}_{jk_j}$ and $\bar{\beta}'_{jk_j}$, $\forall j, k_j$ (a $N(0, \nu^2)$ where $\nu^{-2} = 0.0001$), and (a_m, b_m), (a uniform distribution (0,100000)).

Inferences from this model were derived by obtaining posterior samples using a Markov chain Monte Carlo sampler. All computation was performed using the software package BUGS, Bayesian Inference Using Gibbs Sampling (http://www.mrc-bsu.cam.ac.uk/bugs/). All of the results reported in section 4 are the posterior means obtained from aggregating the draws of three runs of the sampler from different starting points with a burn-in period of 6000 draws, and a total run length of 10,000 draws. Convergence was assessed using the F-test approach of Gelman and Rubin (1992).

3 The Experiment

An experiment was designed to provide a basic demonstration of our model on rating conjoint data with missing attributes. Our interest lies in providing a demonstration of our approach, as well as to begin a preliminary understanding of:

1. Do people use missing attributes, and their levels, to evaluate products?

2. If yes, do they infer missing attribute levels from all information they learn about the product profiles; do they reinforce patterns?

We assume that a consumer has minimal prior information about the product, although we do estimate this. Therefore, we are able to impose a “prior” structure that varied across respondents in a systematic way (described below). First, through a learning process, we create a “prior” for each individual by controlling the products that she sees in a “learning phase”. Second, we ask participants to rate products with (or without, in the control group) missing attributes. A second control group worked on a self-explicated conjoint task to act as a second baseline.

3.1 Stimulus

Digital cameras were selected for this experiment as we wanted a relatively new product category where the frequency of attribute levels and the correlation structure of the attributes are mostly unknown. This would allow us to manipulate the frequencies and impose a prior as desired. From our demographic questions, less than 10% of our subjects owned digital cameras or claimed to have extensive prior expertise. We chose digital cameras with 6 attributes (as the full profile task) as research that we conducted indicated that digital cameras could be described well using 6 features. A summary of the digital camera attributes used are listed in Table 3. In our experimental condition, all attributes are simplified to have two realistic levels (note SLR = single-lens reflex camera, a size bigger than medium). Of course, having only six attributes and two levels per attribute provides only a very limited empirical test of our model, and is an area for much future study.

[Insert Table 3 Here]

3.2 Experiment Groups

The experiment was designed to run on a university network. One hundred and thirty undergraduate students from a large east-coast university participated in the experiment. The subjects were randomly assigned into four groups: self-explicated, full profile (0 missing), one missing attribute, and two missing attributes, with group sizes 17, 23, 47, and 43, respectively. Group size difference is due to different class sizes. Across the conditions, less than 10% own digital cameras; 40% are females and 60% are males.

The experiment was composed of two phases: the learning (prior) phase and the rating phase. In the learning phase, the subjects were provided with information about digital cameras. They were then shown 20 digital camera profiles listed in a single table. We control the consumers' priors by manipulating the digital camera profiles they see in the learning phase. In the rating session, the subjects were asked to rate the attractiveness of different digital cameras (some with partial product profiles depending on the treatment condition).

3.3 Learning Phase

In the learning phase, all subjects were shown 20 digital camera profiles. The priors of the subjects before the rating phase were manipulated by the learning phase profiles. The purpose of this learning phase manipulation is two-fold. First, if the relationship say between price and maximum resolution, as described next, can be influenced by showing subjects profiles of a given structure, then managerial practice would suggest prior manipulations of this type could be valuable (it does happen in real life; for example, as a result of "education" by Intel, most consumers falsely believe that a computer with faster CPU clock speed calculates faster). Secondly, we wanted to test out our model, for a given attribute correlation. As we manipulated the priors between digital camera *price* and *maximum resolution*, we wanted to test if λ_{i6} would impact the "count" for resolution when it is missing (price is never missing as it is an OM attribute).

Each subject was randomly assigned to one of eleven prior coincidence structures representing a different level of coincidence between *price* and *maximum resolution* (while keeping the coincidences between other attributes orthogonal.) Specifically, they were assigned to read a table with a specific coincidence value (between 0 and 10) between price and maximum resolution. For example, a coincidence value of 10 indicates that among the 10 profiles that have low price ($159), all have low resolution (800×600); a value of zero would indicate that among the 10 profiles that have low price ($159), all of them have high resolution (1024×768). Such coincidence structure could potentially affect $\Pr(\boldsymbol{x}'_{ij}(t) = \boldsymbol{l}_{jk_j})$ if the learning phase carries over to the rating phase. That is, we will test empirically the ability to manipulate the rating phase data by co-varying price and maximum resolution at levels $0, 1, ..., 10$ in the learning phase, and then estimating λ_{i6} in our model and seeing its correlation with the subject's prior manipulation.

To ensure that subjects followed and attended to all information in the table of 20 learning profiles provided, they were asked to count five of the pairwise coincidences after they had read the tables. Among the five questions, one of them asked the subjects to count the coincidence between price $159 and maximum resolution 800×600, the manipulated coincidence, while the other four questions were randomly chosen to ask the subjects to count other coincidences. The sequences of these questions were randomized so as not to bias the results. Their responses to these questions suggested that they had paid attention to the coincidence counts.

3.4 Rating Phase

In the rating phase, the design is orthogonal. In the one or two missing attribute group, one or two out of *four* PM attributes are removed from the designed conjoint cards, respectively. We fixed two attributes to be OM as we wanted to see the impact of imputation of missing levels on observed attribute partworths. We utilized a Plackett-Burman Design (Green et al 1978) to create the profile cards. The sequences in which the profile cards were shown were generated randomly and varied across respondent. Each subject saw 24 profiles in the rating phase. Debriefing questions after the experiments provided evidence that the subjects do "notice" that attributes are missing and utilize this fact in their ratings. Response time was recorded, which could be used as a proxy for the difficulty of the task. Each response time was the time (in seconds) measured between successive rating score responses being keyed in. An analysis of the response time data across missing attribute conditions (0, 1, and 2) indicated that respondents in the two missing case, spent considerably less time than in either of the other two cases ($p < 0.01$), corresponding to an average of 31 seconds less across the 24 profile rating tasks. No significant differences were found in response time between the zero and one missing attribute conditions.

4 Results

We utilized the first 20 profiles for each subject to the calibrate model, and the last four as holdout for validation.

Denote row vectors $\boldsymbol{\lambda_i} = [\lambda_{i1}, \lambda_{i2}, ..., \lambda_{i6}]$ (length of $\boldsymbol{\lambda_i}$ varies with different models in Table 4), $\boldsymbol{\beta_i} = [\beta_{ij1}, \beta_{ij2}, ..., \beta_{ij(K_j-1)}]$, and $\boldsymbol{\beta'_i} = [\beta'_{ij1}, \beta'_{ij2}, ..., \beta'_{ij(K_j-1)}]$. Given $\Pr(\boldsymbol{x}'_{ij}(t) = \boldsymbol{l}_{jk_j})$ from Equation (2), the likelihood function is

$$L(\alpha_i, \boldsymbol{\beta_i}, \boldsymbol{\beta'_i}, \boldsymbol{\lambda_i}, \gamma_i, \sigma_i^2; \boldsymbol{y}_i | \boldsymbol{X}_i, \boldsymbol{R}_i) = \prod_{t=1}^{20} \left\{ \sum_{k_j=1}^{K_j} \left[\frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(\Delta_i(t) - \gamma_i \Delta_i(t-1))^2}{2\sigma_i^2} \right) \cdot \Pr\left(\boldsymbol{x}'_{ij}(t) = \boldsymbol{l}_{jk_j}\right) \right] \right\}. \tag{7}$$

That is, we integrate the conjoint regression model with respect to the imputation model, i.e. stick in the considered value for attribute j to person i for all K_j possible values, and weight them by their probability of being the imputed corresponding level.

We estimated a total of six models, with differing degrees of generality. These models are grouped into four categories: (1) prior models, (2) imputation based on missing PM attributes only, (3) imputation based on missing and non-missing PM attributes, and (4) imputation based on the OM attribute (Price), missing and non-missing PM attributes. Table 4 shows these models and their relationships for both the one and two missing attribute cases. The estimation was done using the Bayesian hierarchical structure described in Section 2.4.

Insert [Table 4 Here]

Table 5 shows the relative performance for the six models. We show results for the three extant models (ignore, recency, and averaging), as well as one model in each of the three classes (Models 4, 5, and 6). For each model, we report (1) the marginal log-likelihood as computed by the harmonic mean of the log-likelihood values, and (2) mean absolute errors (MAE), both in sample and out of sample. Using all three measures, our models (Models 4-6) perform better than the prior models (Models 1-3). Specifically, Models 4 through 6 consistently perform better as more information is used for imputation.

[Insert Table 5 Here]

4.1 Imputation Based on Missing PM Attributes

As discussed above, the extant models assume that subjects either ignore the missing attribute, use the most recent occurrence of a missing attribute, or compute the average of all past occurrences. Out of models 1-3, the averaging model (Model 3) performs better in terms of marginal log-likelihood and in- and out-of-sample mean absolute error (MAE). Model 4 relaxes the assumptions of Model 3: it allows a separate λ for each missing PM attribute, which decays geometrically.

Compared to the averaging model, Model 4 performs better in terms of marginal log-likelihood, in sample and out of sample MAE. Such results suggest that the relaxation of allowing for heterogeneous geometric decay helps in terms of model performance and better captures the actual rating process when missing attributes exist.

4.2 Imputation Based on Missing and Non-missing PM Attributes

The above class of model assumes that subjects impute a missing level of an attribute using only information within that PM attribute. A natural extension is to account for covariation from the non-missing PM attributes. In Model 5, we assume each attribute to take a different λ when it is present and when it is missing; however, the value of λ is assumed to be common across all PM attributes when they are non-missing. As indicated in Table 5, Model 5 fits better than Models 1-4 in terms of marginal log-likelihoods, in-sample MAE, and out-of-sample MAE.

4.3 Imputation Based on OM Attribute, Missing, and Non-Missing PM Attributes

To fully test the information used by the subjects when inferring missing attribute levels, in addition to the last sets of models, we add price (an OM attribute) to impute the missing level of maximum resolution. Recall that we manipulated the correlation structure between price and resolution in the learning phase. Models 6 extends Model 5 by allowing price to be used in the imputation process for missing maximum resolution levels. Such a relaxation improves the marginal log-likelihood[10], in-sample, and out-of-sample MAE. The results strongly suggest that subjects do use OM attributes to infer missing attribute levels in a significant way. In fact, Model 6 outperforms Models 1-3 significantly by all the measures we considered in Table 5. Specifically, Model 6 decreases the out-of-sample MAE by 4.1%, 5.3%, and 6.5% over Models 1-3, respectively, in the one-missing attribute case; and 1.6%, 3.0%, and 5.9% over Models 1-3, respectively, in the two-missing attributes case.

A more detailed analysis at the individual-level between the estimated effect λ_{i6} (price and maximum resolution) and the prior manipulated covariation between price and maximum resolution (0,1,...,10) was done in a number of ways. (1) First, we note that λ_{i6} is significantly different from zero (in the population, $p < 0.01$, in both one- and two-missing attributes cases), suggesting significant effects overall. (2) An analysis at the individual-level (without shrinkage), indicated a significant effect (correlation = 0.15, $p < 0.001$ in the one-missing attribute case; correlation = 0.12, $p < 0.001$ in the two-missing attributes case) between the prior manipulation and λ_{i6}. Overall, these findings suggest that the subjects' "priors" could be manipulated to influence the way they infer missing attributes.

The average of λ values of the best-fitted model (Model 6) are provided in Table 6. The esti-

[10] The marginal log-likelihood becomes positive in some cases. A positive log-likelihood occurs when the variance of the normal distribution becomes very small.

mated λ values are different in the one- and two-missing attributes cases, which is not surprising: when different number of attributes are missing, the weights which reflect how information from non-missing attributes is used change correspondingly.

[Insert Table 6 Here]

Notice all the λ values are significantly larger than 0 and smaller than 1, indicating the actual imputation procedure is different from the pure effect of any one of Models 1-3.

4.4 Estimated Part-worths and Priors

Table 7 reports the mean and standard deviation of partworths of Model 6, the best fitting model. $\beta'_{ij1}, (j = 1, ..., 4, \forall i)$ are the partworths of the attributes when they take level 0 and are present (compared to taking level 0 and not being present, but instead being imputed as a zero). $\beta_{ij1} + \beta'_{ij1}, (j = 1, ..., 4, \forall i)$ are the partworths of the attributes when they take level 1 and are present (compared to taking level 0 and not being presented). To compare our results, therefore, with traditional conjoint partworths, we note that with *all attributes present* and the "low" level attributes coded as zero (as is standard), the partworths represent the difference in utility between the high and low attribute levels. To align with our case, the "traditional partworths" from our model are $(\beta_{ij1} + \beta'_{ij1}) - \beta'_{ij1} = \beta_{ij1}, (j = 1, ..., 4, \forall i)$; that is, the effect of being high when shown minus the effect of being low when shown.

As mentioned earlier, we get a "bonus" in that we can assess the effect of imputed versus not imputed in our conjoint design, in addition to level 1 (high) versus level 0 (low). In our model, these are the partworths, β'_{ij}. We find that all elements $\beta'_{ij}, (j = 1, ..., 4, \forall i)$ are significant at 0.05 level, which means that when an attribute is present, it is given significantly greater weight. This finding is consistent with extant research (Johnson 1987, Levin et. al. 1986, Louviere and Johnson 1990, Meyer 1981). Alternatively, we note that the combined tests (present or not, and levels 1 or 0), for each of the attributes, suggest that it is not the attribute level inferred, but the presence of the attribute that influences the weight put on the attribute. We believe this is an area for interesting future study.

Table 8 reports the relative importance of (traditional) partworths, i.e. when high level is shown versus low level is shown, from Model 6, and from the case where there are no missing attributes. While we observe relatively high stability in the rankings (for instance, storage and size are always last, resolution is always most important, and the other three are relatively close in importance), there are changes in the magnitude of the relative importance of the partworths. This finding, that partworths themselves are "biased" (as compared to the full profile condition), is consistent with extant research (Johnson 1987, Levin et. al. 1986, Louviere and Johnson 1990). However, given that we also find that the relative rankings stay fairly stable, there is *prima facia* evidence that similar rating processes are going on. Interestingly, in the two-missing attributes case, when less attribute levels are available for imputation, the more important attributes in the no-missing case become less important, the less important ones become more important. Thus, there is a "regression" effect in partworths when subjects evaluate partial profiles when less information is provided.

We note, that one way to interpret the observed changes in partworths is that consumers "construct" rather than "retrieve" utilities. Since the set of all available information changes with successive profiles, the utilities can change even for identical profiles if they appear at different points in time. This view is not new and has been established by consumer researchers (Bettman and Zins 1977, Payne, Bettman, and Johnson 1992) We shall discuss in Section 5 how this phenomenon opens up an interesting area for managerial action and impact.

[Insert Table 7 Here]

[Insert Table 8 Here]

Finally, we report on the model results with regards to the carry-over effect from one rating's error $\epsilon_{i(t-1)}$ to another, and from the priors $N_{ijk_j}(0|\boldsymbol{l}_{jk_j})$. The AR(1) carry-over effect is statistically significant with $\bar{\gamma} \approx 0.1$ (in both one- and two-missing attributes cases), suggesting that people do anchor somewhat on previous values. This result suggests that the order in which previous profiles are presented could influence the subjects' rating for a current profile.

None of the estimated prior parameters ζ_i $(i = 1, ..., 47)$ and ω_{jk_j} $(j = 1, ..., 6, k_j = 1, 2)$ is significantly different from zero at $p = 0.05$ level in the one-missing attribute case, an indication that the subjects have a weak prior on the product. Consequently, the average values of the experience counts, as shown in Table 9, for Model 6, are typically small. These initial experience counts thus exert some minor influence on the imputation of the early profiles but decay quickly when more profiles are shown. However, some of the prior parameters become significant in the two-missing attributes case. Specifically, ω's for resolution and size, the PM attributes people are probably most familiar with, are fairly significant. This shows that when less information becomes available, people may depend more on their "priors" to make judgements. This certainly requires further study beyond the empirical example provided here.

[Insert Table 9 Here]

5 Conclusions, Caveats, and Future Research

We have developed a learning model to describe how consumers impute missing levels in partial conjoint profiles. In our model, consumers match patterns and develop inferences based on their prior exposures. Our model extends the Averaging and Recency models and shows that consumers may infer missing attribute levels using both missing and non-missing attribute information. We have shown that our best-fit models significantly outperform the prior models both in-sample and out-of-sample.

The Ignore model is inadequate because consumers appear to consider missing attribute levels. Neither the Averaging nor Recency models does significantly better because consumers impute missing attribute levels using prior levels of non-missing attributes (whether PM or OM). At the same time, significant correlation between the manipulated coincidence in the learning phase and the estimated decay parameter provides evidence that the consumer's prior could be influenced by "communication" and "experience". Consequently, managers can influence the overall attractiveness of a product to a consumer by making the consumers learn "prior knowledge" that favors the product.

As we have mentioned earlier, this research has several caveats.

1. Our choice of six attributes, each with two levels, and one or two missing is a nice first step, however, it's a rather simple demonstration in comparison to more realistic empirical settings.

2. As a demonstration of our model, our rating conjoint experiment does not provide empirical evidence for the applicability of our model to pairwise comparison or choice-based conjoint analysis, although theoretically, such application is viable.

We see at least three future research opportunities:

1. One application of the model is to find the trade off between the number of profiles shown and the number of attributes shown in each profile. From an econometric perspective it would be interesting to fix the number of attribute levels shown, and see how different combinations of the number of attributes and profiles, would lead to differing amounts of information.

2. As in Section 2, we assumed here a pattern matching model in which attributes match or do not (0/1). A more general distance model may be more appropriate in which attribute levels have distances from each other. Such machinery is already in the marketers' toolbox as MDS studies are used for such purposes. We therefore believe that two promising areas for future studies would be: (a) to combine conjoint analysis and MDS studies to impute missing attribute levels, and (b) to create a "latent" perceptual mapping model for missing attribute levels in conjoint.

3. As mentioned previously, and shown in earlier research work, the missingness of attributes may indeed change the relative importance of attributes. While our work additionally shows that, and for the most important attributes, whether this is true generally is unclear, and what may moderate this effect may also be of interest. If one were to use this for managerial action, it would be necessary to conduct future carefully designed studies to assess this.

References

[1] Bettman, J. R., and M. Zins (1977), "Constructive Processes in Consumer Choice," *Journal of Consumer Research*, 4, 75-85

[2] Bickart, B. A. (1993), "Carryover and Backfire Effects in Marketing Research", *Journal of Marketing Research*, 30 (February), 52-62.

[3] Broniarczyk, S. M. and J. W. Alba (1994), "The Role of Consumers' Intuitions in Inference Making," *Journal of Consumer Research*, Vol. 21 (December), 393-407

[4] BUGS: Bayesian Inference Using Gibbs Sampling (1996), Version 0.5, Spiegelhalter, D.J., A. Thomas, N. G. Best, and W. R. Gilks

[5] Camerer, C. and T.-H. Ho (1999), "Experience-Weighted Attraction Learning in Normal Form Games," *Econometrica*, Vol. 67, No. 4 (July), 827-874

[6] Feldman, J. M. and J. G. Lynch, Jr. (1988), "Self-generated Validity and Other Effects of Measurement on Belief, Attitude, Intention, and Behavior," *Journal of Applied Psychology*, 73 (August), 421-435

[7] Gaul, W. (1998), "Linking Conjoint Analysis and QFD," Unpublished Dessertation, Universitt Karlstuhe

[8] Gelfand, A. E., and A. F. M. Smith (1990), "Sampling-Based Approaches to Computing Marginal Densities", *Journal of the American Statistical Association*, 85, 398-409.

[9] Gelman, A., and D. B. Rubin (1992), "Inference from Iterative Simulation Using Multiple Sequences", *Statistical Science*, Vol. 7, 457-511.

[10] Green, P. E. (1974), "On the Design of Choice Experiments Involving Multifactor Alternatives," *Journal of Consumer Research*, Vol. 1, 61-68

[11] ——, J. D. Carroll, and F. J. Carmone (1978), "Some New Types of Fractional Fractorial Designs for Marketing Experiments," *Research in Marketing*, Vol. 1, 99-122

[12] —— (1984), "Hybrid Model for Conjoint Analysis: An Expository Review," *Journal of Marketing Research*, Vol. 21, 155-169

[13] Hoch, S. J., E. T. Bradlow, and B. Wansink (1999), "The Variety of an Assortment," *Marketing Science*, Vol. 18, No. 4, 527-546

[14] Huber, J. and J. W. McCann (1982), "The Impact of Inferential Beliefs on Product Evaluations," *Journal of Marketing Research*, 19 (August), 324-333

[15] Johnson, R. D. (1987), "Making Judgments When Information is Missing: Inferences, Biases, and Framing Effects", *Acta Psychologica*, Vol. 66, 69-82

[16] —— and I. P. Levin (1985), "More than Meets the Eye: The Effect of Missing Information on Purchase Evaluations," *Journal of Consumer Research*, 12 (September), 169-177

[17] Levin, I. P., R. D. Johnson, Patricia J. Deldin, Laura M. Carstens, LuAnne J. Cressey, and Charles R. Davis (1986), "Framing Effects in Decisions with Completely and Incompletely Described Alternatives", *Organizational Behavior and Human Decision Processes*, Vol. 38, 48-65

[18] Little, R. J. A. and D. B. Rubin (1987), Statistical Analysis with Missing Data, New York: John Wiley & Sons, Inc.

[19] Louviere, J. J. and R. D. Johnson (1990), "Reliability and Validity of the Brand-Anchored Conjoint Approach to Measuring Retailer Images", *Journal of Retailing*, Vol. 66, No. 4 (Winter), 359-381.

[20] Lynch, J. G., Jr. and T. K. Srull (1982), "Memory and Attentional Factors in Consumer Choice: Concepts and Research Methods," *Journal of Consumer Research*, Vol. 9, June, 18-37

[21] Meyer, R. (1981), "A Model of Multiattribute Judements Under Attribute Uncertainty and Informational Constraint," *Journal of Consumer Research*, Vol. XVIII (November), 428-41

[22] Payne, J. W., J. R. Bettman, and E. J. Johnson (1992), "Behavioral Decision Research: A Constructive Processing Perspective," *Annual Review of Psychology,* Vol. 43, 87-131

[23] Plackett, R. L. and J. P. Burman (1946), "The Design of Optimum Multifactorial Experiments", *Biometrika*, Vol. 33, 305-325.

[24] Ross, W. T., Jr. and E. H. Creyer (1992), "Making Inferences about Missing Information: The Effects of Existing Information," *Journal of Consumer Research*, 19 (June), 14-25

[25] Schuman, J. and S. Presser (1981), "Questions and Answers in Attitude Surveys, Experiments in Question Form, Wording, and Context", New York: Academic Press.

[26] Troutman, C. M. and J. Shanteau (1977), "Inferences Based on Nondiagnostic Information," *Organizational Behavior and Human Performance*, June

[27] Tversky, A. and D. Kahneman (1974), "Judgment Under Uncertainty: Heuristics and Biases," *Science*, 185, 1124-1131.

[28] West, P. M., C. L. Brown, and S. J. Hoch (1996), "Consumption Vocabulary and Preference Formation," *Journal of Consumer Research*, 23(2), 120-135.

[29] Wilson, T. D., C. E. Houston, K. M. Etling, and N. Brekke (1996), "A New Look at Anchoring Effects: Basic Anchoring and Its Antecedents," *Journal of Experimental Psychology, General*, 125(4).

[30] Yamagishi, T., and C. T. Hill (1981), "Adding versus Averaging Models Revisited: A test of a Path-Analytic Integration Model," *Journal of Personality and Social Psychology*, 41:13-25

Time	OM	PM		
	Attribute 1 (A_1)	Attribute 2 (A_2)	Attribute 3 (A_3)	Attribute 4 (A_4)
1	1	0	?	1
2	0	1	1	?
3	1	?	1	0

Table 1. An illustrative example

		Attribute Level ($K_j = 2$)	
		High: $x'_{ij1}(t) = 1$	Low: $x'_{ij1}(t) = 0$
Attribute Shown?	Yes	$\beta_{ij1} + \beta'_{ij1}$	β'_{ij1}
	No	β_{ij1}	0

Table 2. Partworths of the same attribute

Attribute	**Delay Between Shots**	**Storage Media**	**Maximum Resolution**	**Camera Size**	**Price**	**Mini Movie**
Level 0	4 Seconds	Floppy Disk	800X600	SLR	$239	No
Level 1	2 Seconds	Removable Memory	1024X768	Medium	$159	Yes

Table 3. Digital camera attributes

Model Category	**Model**	**Individual level # of parameters**	**λ Values of PM Attributes**					**λ Values of OM Attributes**	
			Missing PM				**Non-Missing PM**		
			Delay	**Storage**	**Resolution**	**Size**		**Price**	**Mini-Movie**
Prior Models	1 §	12	---				---	---	---
	2 †	12	$\lambda_1=\lambda_2=\lambda_3=\lambda_4=\lambda, \to 0$				---	---	---
	3 ‡	12	1				---	---	---
Imputed based on missing PM	4	14	λ_1	λ_2	λ_3	λ_4	---	---	---
Imputed based on missing and non-missing PM	5	17	λ_1	λ_2	λ_3	λ_4	λ_5	---	---
Imputed based on OM & missing and non-missing PM	6	18	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6 *	---

§ Ignore missing model
† Recency Model
‡ Averaging Model
* λ values are used for price-resolution imputation only

Table 4. Description of models

Model	1-missing			2-missing		
	Marginal Log-likelihood	Mean absolute error		Marginal Log-likelihood	Mean absolute error	
		In-sample	Out-of-sample		In-sample	Out-of-sample
1	-420.7	0.807	1.366	-441.7	0.861	1.382
2	-380.8	0.778	1.386	-396.9	0.823	1.390
3	-113.5	0.733	1.314	105.7	0.500	1.381
4	-40.3	0.694	1.310	210.4	0.461	1.360
5	-20.0	0.602	1.292	277.0	0.459	1.341
6	81.6	0.580	1.281	434.0	0.434	1.300

Table 5. Performance of different models

Model	Missing PM				Non-missing PM	OM (price)
	Delay	Storage	Resolution	Size		
1-missing	0.306	0.095	0.829	0.240	0.977	0.247
2-missing	0.981	0.429	0.032	0.070	0.697	0.743

Table 6. Average of the λ's of the best-fitted model (Model 6)

Model	Coefficients		Delay	Storage	Resolution	Size	Price	Mini-Movie	Intercept
1-missing	β_{ij1}	Mean	1.328	0.046	1.384	0.092	0.829	1.296	0.592 (0.426)
		SD	0.124	0.182	0.133	0.144	0.124	0.158	
	β'_{ij1}	Mean	0.844	0.745	0.876	0.737	---	---	
		SD	0.117	0.145	0.124	0.136	---	---	
2-missing	β_{ij1}	Mean	1.001	0.565	1.267	0.469	1.000	1.085	0.526 (0.140)
		SD	0.004	0.177	0.088	0.150	0.004	0.130	
	β'_{ij1}	Mean	1.196	1.238	1.168	1.119	---	---	
		SD	0.065	0.106	0.063	0.063	---	---	

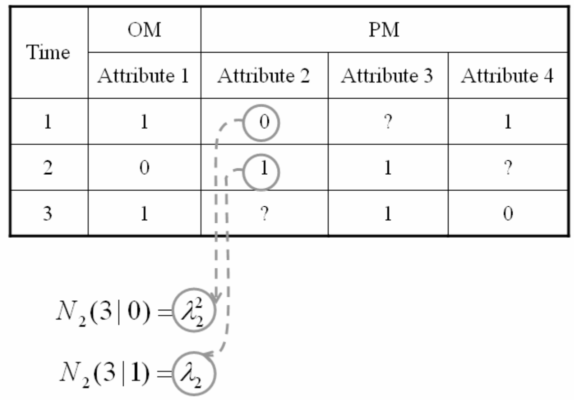
Table 7. Average and standard deviation of the partworths of the best-fitted model (Model 6)

Model	Delay	Storage	Resolution	Size	Price	Mini-Movie
0-missing case	0.126	0.029	0.317	0.056	0.198	0.275
1-missing case, Model 6 (β_{ij1})	0.267	0.009	0.278	0.018	0.167	0.261
2-missing case, Model 6 (β_{ij1})	0.186	0.105	0.235	0.087	0.186	0.201

Table 8. Comparison of relative importance of partworths

Models	1-missing		2-missing	
Levels	0	1	0	1
Delay	1.025	0.931	4.198	2.461
Storage	0.194	0.722	2.898	4.812
Resolution	0.422	0.539	10.068	16.816
Size	0.453	0.441	4.439	4.858
Price	0.967	1.004	2.062	2.006
Mini-Movie	1.005	1.017	2.002	2.022

Table 9. Average of estimated priors $N_{ijk_j}(0 \mid \cdot)$

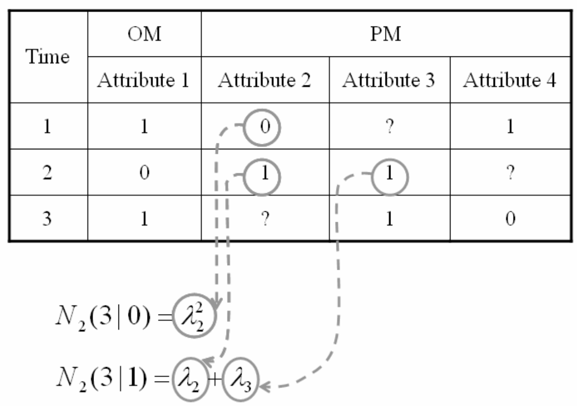


Time	OM	PM		
	Attribute 1	Attribute 2	Attribute 3	Attribute 4
1	1	0	?	1
2	0	1	1	?
3	1	?	1	0

$N_2(3|0) = \lambda_2^2$

$N_2(3|1) = \lambda_2$

Figure 1A. Imputing from missing PM only

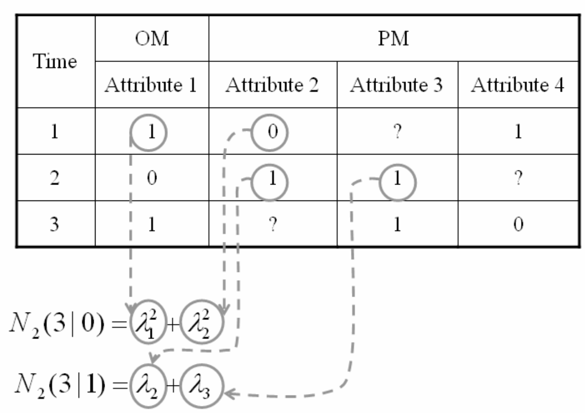


Time	OM	PM		
	Attribute 1	Attribute 2	Attribute 3	Attribute 4
1	1	0	?	1
2	0	1	1	?
3	1	?	1	0

$N_2(3|0) = \lambda_2^2$

$N_2(3|1) = \lambda_2 + \lambda_3$

Figure 1B. Imputing from missing and non-missing PM



Time	OM	PM		
	Attribute 1	Attribute 2	Attribute 3	Attribute 4
1	1	0	?	1
2	0	1	1	?
3	1	?	1	0

$N_2(3|0) = \lambda_1^2 + \lambda_2^2$

$N_2(3|1) = \lambda_2 + \lambda_3$

Figure 1C. Imputing from OM, missing and non-missing PM