Seoul Journal of Business
Volume 19, Number 1 (June 2013)

Dynamic Heterogeneous Choice Heuristics: A Bayesian Hidden Markov Mixture Model Approach

JIN GYO KIM*
Seoul National University
Seoul, Korea

Abstract

Standard choice models implicitly assume that consumers, in order to maximize their expected utilities, compare each of the alternatives in their choice sets in terms of all available attributes. Consumer-level utility functions are frequently taken as linear, and overwhelmingly so as compensatory. However, due to limitations in information process capacity, characteristics of choice task environment and other internal or external constraints, consumers may search for satisfying alternatives rather than optimal ones by invoking other non-compensatory heuristics which free them from arduous attribute-by-attribute comparison. The question arises as to how often these non-compensatory rules are applied, and whether researchers can detect them using only standard data sources.

This study aims to address two main issues regarding consumers' use of decision-rules and heuristics in the real world: (1) whether they are heterogeneous across consumers and (2) whether they are changing for individual consumers over time. To these ends, we extend the standard linear compensatory rule assumption to more faithfully capture dynamic heuristic usage for each consumer. There are three reference heuristics studied in this paper, the well-known linear compensatory, disjunctive and conjunctive rules. Conditional on this known set of possible heuristics, a dynamic heterogeneous hidden Markov mixture choice model is developed to capture heuristic dynamics at the individual-level. When estimated on detergent scanner data, the proposed model offers strong evidence supporting both heterogeneity and dynamics in heuristics usage.

Keywords: Choice Heuristics, Choice Models, Hidden Markov Model, Heterogeneity, Dynamics, Bayesian Methods, Markov chain Monte Carlo

---

* Jin Gyo Kim, Graduate School of Business, Seoul National University, Shillim-Dong, Kwanak-Ku, Seoul, 151-916, Korea. Correspondence regarding this manuscript should be addressed to the author by e-mail to jingkim@snu.ac.kr, or by surface mail to the above address.

1. INTRODUCTION

Standard random utility models of consumer choice, with few exceptions, assume that compensatory rules capture the comparison process for multiattribute alternatives. Underlying this presumption, though seldom stated as such, is the belief that to make a choice requires that all alternatives in one's choice set are compared on all relevant attributes. Compensatory rules simply require as much, since good results on one valued attribute can offset deficiencies in another. By far the most popular and widely-applied of the compensatory rules is the linear-additive, simply a weighted sum of attributes which can presumably be easily compared across potential choices to arrive at the one offering 'maximal expected utility'.

There is, however, a large literature showing that choices in the real world are often sub-optimal, at least compared with what one might have chosen with limitless information, time and cognitive resources. Subject to limitations in processing capacity, consumers may use seemingly sub-optimal satisfying rules by searching through a set of alternatives until a satisfactory one is found (cf., Simon 1975). For example, Tversky (1972) maintained that unsatisfactory alternatives are eliminated by sequentially (and probabilisitically) picking aspects until only one alternative remains. This EBA-based rule is intrinsically non-compensatory: once an alternative is eliminated, how strong it might have been on remaining attribute comparisons is irrelevant. An appealing feature of such heuristics or rules-of-thumb is that they can free consumers from long sequences of complex calculations. The savings in time and effort are offset, of course, by the decrement in the quality or utility of the chosen item. Simply put, non-compensatory heuristics offer simplicity at the potential cost of optimality.

Previous studies have proposed various types of heuristics, among them linear compensatory, disjunctive, conjunctive, lexicographic, elimination-by-aspects, lexicographic semiorder, min-max and max-min rules (cf., Bettman 1979; Wright 1975). An interesting empirical question is which of these choice heuristics is most widely used by consumers. Most previous behavioral studies on heuristics were conducted in a laboratory setting, and so could not provide conclusive evidence about consumers' heuristics usage in the real world. Other disciplines such as statistical or econometric modeling have also been nearly silent on the question, instead using the standard linear-compensatory rule in the lion's share of empirical field studies.

However, several studies in empirical modeling did attempt to deviate from the assumption of compensatory heuristics. Most of these studies were, however, two-stage choice models (cf., Gensch 1987; Manrai and Sinha 1989; Roberts and Lattin 1991): (1) a screening stage for the formation of consideration sets and (2) a choice decision stage given consideration sets. The second choice stage was typically assumed to follow compensatory choice heuristics. Other studies also introduced parameters to capture probabilities of alternatives passing the first screening stage (Chiang *et al.* 1999; Swait 1987; Gaudry and Dagenais 1979; Siddarth *et al.* 1995), although these studies did not incorporate non-compensatory heuristics for the first screening stage explicitly. Only a few studies explicitly introduced non-compensatory heuristics at the first screening stage: a feature-based elimination process similar to elimination-by-aspect heuristics (Andrews and Manrai 1998; Gilbride and Allenby 2006) and conjunctive/disjunctive heuristics (Gilbride and Allenby 2004). However, these two studies could not fully address consumers' heuristics usage, since a hybrid decision process – non-compensatory heuristics followed by compensatory ones – was assumed *a priori*; such presumed hybrid choice decision processes preclude other possibilities, for example, that there are some consumers who do not use non-compensatory rules, or that compensatory rules precede non-comepensatory rules for some consumers. To address real-world heuristics usage, it seems important to capture heterogeneity of heuristics usage across consumers without such presumed hybrid choice-decision processes.

Furthermore, it is also possible that consumers' heuristics are not constant, but vary with experience. One can identify at least two reasons why heuristics might change over time. The learning hypothesis suggests that the realization of outcomes of previously chosen alternatives through consumption provides consumers with knowledge about the adequacy of the choice heuristics previously used (Bettman 1979). If a choice heuristic leads to an unsatisfying outcome, consumers may shift to other, different heuristics on the next choice task. An second possible explanation involves characteristics of the external choice environment. For example, Wright (1975) found that compensatory rules are likely to be used less frequently as the number of alternatives increases. Payne (1982) also found that simplifying heuristics are preferred under time pressure in complex decision environments.

This paper aims to develop a comprehensive model of individual-level, dynamic consumer heuristics usage in real shopping environments, using only standard data sources. Specifically, we address two major issues: (1) whether heuristic usage is heterogeneous across consumers and (2) whether heuristics themselves are dynamic at the individual-level. To do so, we relax the standard linear compensatory rule assumption, and propose a heterogeneous dynamic hidden Markov mixture choice model.

The choice literature is replete with potential heuristics among which a consumer might choose. While the proposed methodology can readily accommodate a wide variety of heuristics, here we use just three most common reference heuristics (as did Gilbride and Allenby (2004)):

1. Linear compensatory rule: All alternatives are compared in terms of all attributes, using attribute-specific importance weights. The important feature is that positive and negative deviations among attributes can balance or compensate for one another, in this case in a linear fashion. Linear compensatory rule therefore requires substantial computational processing of trade-offs among attributes. This choice heuristics has been 'gold standard' in random utility theory.
2. Disjunctive rule: A non-compensatory heuristic wherein consumers are assumed to have minimal standards, or cutoffs, for each attribute. Acceptable alternatives are ones that pass *at least one* such cutoff. If there are multiple acceptable alternatives, it is not clear which of these alternatives is chosen. The disjunctive rule is commonly invoked when the chosen product must be greatly superior along one dimension.
3. Conjunctive rule: A non-compensatory heuristic under which, like in the disjunctive rule, consumers are assumed to have minimum attribute cutoffs (which may be different from those in the disjunctive rule). However, acceptable alternatives must surpass *all* cutoffs, not merely one. Conjunctive rules are common for multi-attribute decisions where choices must be

'good enough' along every relevant dimension.

First, each of these three heuristics is formulated so that all unknown parameters (e.g., consumer-specific weights for explanatory variables in the linear compensatory rule, and consumer-specific cutoffs in the conjunctive and the disjunctive rules) are fully heterogeneous. Then, a heterogeneous mixing distribution is introduced to estimate consumer-specific probabilities across the three heuristics. Finally, a hidden Markov structure is used to capture the dynamic heuristics usage for each consumer.

This model will allow us to settle several outstanding issues in the choice literature, among them: (1) Which of these heuristics is most widely used by consumers in real-world choice scenarios?; (2) Do consumers differ in terms of the heuristics they employ (in the same product class)?; and, finally (3) Do consumers tend to stick to a particular type of heuristic, or do they switch among them over time?

2. MODEL SPECIFICATION

In this section we describe a model designed to capture heuristics dynamics at the individual consumer level. Throughout, we use three generic subscripts: h denotes a household (h = 1, …, H), j denotes a brand (j = 1, …, J) and t denotes a purchase occasion (t = 1, …, T_h).

Let $y_{ht} = j$ denote the event that household h purchases brand j on its purchase occasion t. Let x_{hjt} denote a K-dimensional covariate vector for household h, brand j and purchase occasion t. Suppose that the covariate vector $x_{hjt} = (\hat{x}'_{hjt}; \tilde{x}'_{hjt})'$ consists of two sub-vectors: (1) a M-dimensional vector of covariates that cutoff levels can be defined, $\tilde{x}_{hjt}$, and (2) a $(K - M)$-dimensional vector of remaining covariates, $\hat{x}_{hjt}$. Let z_h denote a M-dimensional vector of cutoff levels for household h corresponding to $\tilde{x}_{hjt}$.

For non-compensatory choice rules, let $q_{hjt,m}$ (m = 1, …, M) denote an latent indicator variable for the event that a m-th covariate for a brand, $\tilde{x}_{hjt,m}$ passes the corresponding cutoff level, $z_{h,m}$:

$$q_{hjt,m} = \begin{cases} I_{\tilde{x}_{hjt,m} \succeq z_{hm}}, \text{ when } \tilde{x}_{hjt,m} \text{ is a continuous variable,} \\ I_{\tilde{x}_{hjt,m} = z_{hm}}, \text{ when } \tilde{x}_{hjt,m} \text{ is a discrete variable,} \end{cases} \tag{1}$$

where I_A is the usual indicator function for the event A.

In addition, let s_{ht} denote a *latent* discrete random variable indicating one of choice heuristics:

$$s_{ht} = \begin{cases} 1, \text{ if linear compensatory rule is used,} \\ 2, \text{ if disjunctive rule is used,} \\ 3, \text{ if conjunctive rule is used.} \end{cases} \tag{2}$$

Finally, we model the three choice heuristics as follows:

1. Linear additive compensatory rule

$$p(y_{ht} = j \mid s_{ht} = 1, \beta_h) = \frac{\exp(\beta_h' x_{hjt})}{\sum_{i=1}^{J} \exp(\beta_h' x_{hit})}, \tag{3}$$

where β_h is the regression coefficient vector corresponding to x_{hjt}.

2. Disjunctive rule

$$\begin{aligned} &p(y_{ht} = j \mid s_{ht} = 2, z_h^{(2)}) = \begin{cases} \frac{1}{N_{ht}^{(2)}}, \text{ if } j \in E_{ht}^{(2)}, \\ 0, \text{otherwise,} \end{cases} \\ &N_{ht}^{(2)} = \text{the number of brands in } E_{ht}^{(2)}, \\ &E_{ht}^{(2)} = \{ j \mid q_{hjt,m} = 1 \text{ for } \exists m, m = 1, \ldots, M \}, \end{aligned} \tag{4}$$

where $z_h^{(2)}$ is a vector of cutoffs specific to the disjunctive rule.

3. Conjunctive rule

$$\begin{aligned} &p(y_{ht} = j \mid s_{ht} = 3, z_h^{(3)}) = \begin{cases} \frac{1}{N_{ht}^{(3)}}, \text{ if } j \in E_{ht}^{(3)}, \\ 0, \text{otherwise,} \end{cases}, \\ &N_{ht}^{(3)} = \text{the number of brands in } E_{ht}^{(3)}, \\ &E_{ht}^{(3)} = \{ j \mid q_{hjt,m} = 1 \text{ for } \forall m, m = 1, \ldots, M \}, \end{aligned} \tag{5}$$

where $z_h^{(3)}$ is a vector of cutoffs specific to the conjunctive rule.

In (4) and (5), we assume one of the brands in E_{ht}^i, i = 2 and 3, is uniformly chosen if E_{ht}^i is not a singleton set. This assumption is made not only because there does not exist any sound theories guiding which of brands in E_{ht}^i is chosen but also because any brands of E_{ht}^i can be chosen by definition. Gilbride and Allenby (2004) assumed that compensatory rule would be employed to make choices among brands of E_{ht}^i. Aside from that their assumption was made without any theoretical justifications, the compensatory processing of brands in E_{ht}^i are problematic in some cases, for example, where all brands have exactly same values for all or some covariates, or E_{ht}^i is a singleton set. Furthermore, their assumption suffers from a redundancy problem in a sense that covariates employed for the construction of E_{ht}^i are used again for making comparisons between brands. Covariates employed for the first screening stage may, in fact, be irrelevant to choices among brands of E_{ht}^i since all brands of E_{ht}^i cannot be differentiated in terms of their attractiveness by definition. Following Bettman's (1975) argument, we also allow that cutoff levels can differ for the disjunctive and conjunctive rules, an issue we will examine empirically.

Given these three heuristics, this study will investigate the heterogeneity and the dynamics in heuristics usage. The three reference heuristics were chosen largely because they are the most widely studied in the choice literature. Even though they are certainly not the only such and adding others is readily accomplished, we feel that these three may provide sufficient setting for investigating dynamics and heterogeneity in heuristics usage.

2.1 Modeling Changes in Heuristics over Time

Households are allowed to adopt different heuristics across their purchase occasions. In order to capture changes in heuristics over time, let us assume that the *latent* finite state random variable – we will never observe it directly – $s_{ht} \in \{1, 2, 3\}$ evolves according to a Markov process:

$$s_{ht} \mid s_{h,t-1} \sim Markov(\Delta, \pi_1), \tag{6}$$

where π_1 is the initial probability distribution at the first purchase occasion and $\Delta = \{\delta_{mn}\}$ is a 3 × 3 one-step transition probability matrix of the chain, i.e., $\delta_{mn} = \Pr(s_{ht} = n \mid s_{h,t-1} = m)$ with $\sum_{n=1}^{3}\delta_{mn} = 1$:

$s_{h,t-1}$	s_{ht}		
	1	2	3
1	δ_{11}	δ_{12}	δ_{13}
2	δ_{21}	δ_{22}	δ_{23}
3	δ_{31}	δ_{32}	δ_{33}

For identifiability, we invoke the standard assumption the chain is time-homogeneous, irreducible, and aperiodic for all households. Let $\phi_h = \{\phi_h^{(1)}, \phi_h^{(2)}, \phi_h^{(3)}\}$ denote a set of parameters for heuristics i, i = 1, 2, 3, that is, $\phi_h^{(1)} = \{\beta_h\}, \phi_h^{(2)} = \{z_h^{(2)}\}$, and $\phi_h^{(3)} = \{z_h^{(3)}\}$. Then, from (3), (4), (5), and (6), the choice probability follows a finite dynamic hidden Markov mixture distribution:

$$\Pr(y_{ht} = j \mid s_{h,t}, \phi_h) = \begin{cases} \sum_{i=1}^{3} p(y_{ht} = j \mid s_{ht} = i, \phi_h^{(i)})\pi(s_{ht} = i), t = 1, \\ \sum_{i=1}^{3} p(y_{ht} = j \mid s_{ht} = i, \phi_h^{(i)})p(s_{ht} = i \mid s_{h,t-1}), t > 1. \end{cases} \quad (7)$$

2.2 Competing Models

Depending on the assumptions for π_1 and Δ, we have several competing models. Define $\delta_i = (\delta_{i1}, \delta_{i2}, \delta_{i3})$, i = 1, 2, 3. Table 0 lists five such models and their imposed assumptions.

Comparisons among models M_1, M_2 and M_3 allows one to test which of the three reference heuristics is most preferred under

Table 1. List of comepeting models

Model	π_1	Δ
M_1: Linear compensatory rule	(1, 0, 0)	δ_1 = (1, 0, 0); $\delta_2 = \delta_3 = 0_3$
M_2: Disjunctive rule	(0, 1, 0)	$\delta_1 = 0_3$; δ_2 = (0, 1, 0); $\delta_3 = 0_3$
M_3: Conjunctive rule	(0, 0, 1)	$\delta_1 = 0_3$; $\delta_2 = 0_3$; δ_3 = (0, 0, 1)
M_4: Static mixture model	NR*	δ_1 = (1, 0, 0); δ_2 = (0, 1, 0); δ_3 = (0, 0, 1)
M_5: Dynamic mixture model	NR	NR

Note: *: no restriction

the *homogeneous* heuristics assumption. Model M_4 allows one to examine the *heterogeneity* in heuristics across households, assuming that they do not change over time. Finally, model M_5 captures both heuristic *heterogeneity* and *dynamics*. In this way, $M_1 - M_5$ suffice to discern among competing behavioral theories of heuristic usage, using no more than observed, consumer-level choices.

2.3 Modeling Parameters Associated with Heuristics

All parameters associated with heuristics, ϕ_h, are assumed to be heterogeneous across households. For the linear compensatory rule, we model β_h as multivariate normal random-effects:

$$\beta_h \sim N_K(\mu_\beta, \Sigma_\beta), \tag{8}$$

a K-variate normal distribution with a mean vector μ_β and a covariance matrix Σ_β.

Let n_d and n_c denote the number of discrete and continuous variables in z_h, respectively, where $M = n_d + n_c$. Then, for cutoffs, $z_h^{(i)}$, $i = 2, 3$,

1. $\Pr(z_{hn}^{(i)} = r_{nq} \mid \theta_{nq}^{(i)}) = \theta_{nq}^{(i)}, n = 1, \ldots, n_d$, $q = 1, \ldots, Q_n$, when $z_{hn}^{(i)} \in \{r_{n1}, \ldots, r_{n,Q_n}\}$ is a discrete variable,
2. $\Pr(z_{hn}^{(i)} \mid \xi_n^{(i)}) = \eta(z_{hn}^{(i)} \mid \xi_n^{(i)}), n = n_d + 1, \ldots, M$, when $z_{h,n}^{(i)}$ is a continuous variable. The function $\eta(z_{hn}^{(i)} \mid \xi_n^{(i)})$ denotes a probability distribution with parameter $\xi_n^{(i)}$. While any continuous probability distribution functions can be used for $\eta(\bullet)$, in this paper we assume:

$$\Pr(z_{hn}^{(i)} \mid \mu_n^{(i)}, \sigma_n^{(i)}) = N(z_{hn}^{(i)} \mid \mu_n^{(i)}, \sigma_n^{(i)}), \tag{9}$$

a univariate normal distribution with mean $\eta_n^{(i)}$ and variance $\sigma_n^{(i)}$ with $\xi_n^{(i)} = (\mu_n^{(i)}, \sigma_n^{(i)})$.

Define $\Psi^{(i)} = \{\theta^{(i)}, \xi^{(i)}\}$, $i = 2, 3$, where $\theta^{(i)} = \{\theta_1^{(i)}, \ldots, \theta_{n_d}^{(i)}\}$, $\theta_n^{(i)} = \{\theta_{n1}^{(i)}, \ldots, \theta_{nQ_n}^{(i)}\}$ and $\xi^{(i)} = \{\xi_1^{(i)}, \ldots, \xi_{n_c}^{(i)}\}$. Then, rewrite $\Pr(z_{hn}^{(i)} = r_{nq} \mid \theta_{nq}^{(i)})$ and $\Pr(z_{hn}^{(i)} \mid \mu_n^{(i)}, \sigma_n^{(i)})$ together as

$$p(z_h^{(i)} \mid \Psi^{(i)}), i = 2, 3. \tag{10}$$

2.4 Prior

The model specification is completed with the following priors:

$$\pi_1 \sim Dir(l_1, l_2, l_3) \text{ for } \forall h \tag{11}$$

$$(\delta_{n1}, \delta_{n2}, \delta_{n3}) \sim Dir(d_{n1}, d_{n2}, d_{n3}), n = 1, 2, 3, \text{ for } \forall h \tag{12}$$

$$\mu_\beta \sim N_K(f_\beta, V_\beta), \tag{13}$$

$$\Sigma_\beta \sim IW_K(\upsilon_\beta, S_\beta), \tag{14}$$

$$\theta_n^{(i)} \sim Dir(e_{n1}, \dots, e_{n,Q_n}), n = 1, \dots, n_d \text{ and } i = 2, 3, \tag{15}$$

$$\mu_n^{(i)} \sim N(f_n, \upsilon_n), n = n_d + 1, \dots, M \text{ and } i = 2, 3, \tag{16}$$

$$\sigma_n^{(i)} \sim IG(l_n, g_n), n = n_d + 1, \dots, M \text{ and } i = 2, 3, \tag{17}$$

Here $Dir(a_1, \dots, a_n)$ is a Dirichlet distribution with parameters $a_1, \dots, a_n$. The expression $\Sigma \sim IW_K(\upsilon, S)$ denotes that Σ has a K-dimensional inverted Wishart distribution with parameters υ and S, where $\upsilon > 0$ and S is non-singular, that is, $p(\Sigma \mid \upsilon, S) \propto \left|\Sigma\right|^{-(\frac{1}{2}\upsilon + K)} \exp(-\frac{1}{2} \text{tr } \Sigma^{-1}S)$. In addition, the expression $\sigma \sim IG(l, g)$ denotes an inverse gamma distribution with shape parameter l and scale parameter g, that is, $(\sigma \mid l, g) \propto \sigma^{-(l+1)} e^{-g/\sigma}$. Note that all parameters for priors listed above are *known* values.

3 EMPIRICAL APPLICATION

3.1 Data and Independent Variables

The proposed model was fitted to an A. C. Nielson liquid detergent scanner data set for four brands over 96 weeks. Liquid detergent data provide a particularly stringent arena for the detection of choice heuristic heterogeneity and dynamics, given that the category is not characterized by high differentiation or great variation amongst usage occasions. The data consisted of 492 households that made purchases at least six times during the 96 week period. The total number of purchase observations was 6682, and summaries of the data are given in Table 2.

For the linear compensatory rule, we introduced three brand

Table 2. Marketing activities of brands in the liquid detergent data

	Brand A	Brand B	Brand C	Brand D
Choice share	30.0%	28.4%	18.9%	22.8%
Proportion of observations with feature advertising	23.3%	21.7%	6.4%	11.8%
Proportion of observations with display	14.8%	13.5%	6.8%	9.5%
Average and standard deviation of net paid price	\$5.09 (0.58)	\$5.95 (0.64)	\$5.93 (0.51)	\$5.42 (0.68)

dummy variables and three marketing-mix variables. To ensure identifiability, the regression coefficient of the fourth brand was fixed to zero for all households. The predictor variables x_{hjt} for brand j and household h at purchase occasion t were:

$$x_{hjt} = \begin{bmatrix} \text{dummy variable for brand A} \\ \text{dummy variable for brand B} \\ \text{dummy variable for brand C} \\ \text{dummy variable for feature advertising} \\ \text{dummy variable for display} \\ \text{net - paid price} \end{bmatrix}.$$

For the conjunctive and disjunctive rules, there were $M = 4$ attributes for cutoffs as given in Table 3. Note that $z_{h2}^{(i)}$ and $z_{h3}^{(i)}$, $i = 2, 3$, are defined as binary variables. For example, if $z_{h2}^{(i)} = 0$, households can be thought of as indifferent to feature advertising. The cut-off prices are assumed to be bounded in an interval of [*min* (observed prices), *max* (observed prices)].

Finally, the chosen values for the priors are

$$l_n = 1, n = 1, 2, 3,$$

$$d_{n1} = d_{n2} = d_{n3} = 1, n = 1, 2, 3$$

$$f_\beta = 0_6, V_\beta = 30I_6, v_\beta = 2, S_\beta = 30I_6,$$

$$e_{1n} = 1, n = 1, \dots, 15, \text{ for } z_{h1}^{(i)}, i = 2, 3,$$

$$e_{2n} = e_{3n} = 1, n = 1, 2, \text{ for } z_{h2}^{(i)} \text{ and for } z_{h3}^{(i)}, \ i = 2, 3,$$

$$f_4 = 5.5, v_4 = 10, l_4 = 3, g_4 = 10, \text{ for } z_{h4}^{(i)}, \ i = 2, 3.$$

Table 3. Variables for cutoffs in the disjunctive and the conjunctive rules

Cutoffs	Variable	Discrete/ Continuous	Values
$z_{h1}^{(2)}$, $z_{h1}^{(3)}$	Brands preferred	Discrete (Q_1 = 15)	1(A), 2(B), 3(C), 4(D), 5(AB), 6(AC), 7(AD), 8(BC), 9(BD), 10(CD), 11(ABC), 12(ABD), 13(ACD), 14(BCD), 15(ABCD)
$z_{h2}^{(2)}$, $z_{h2}^{(3)}$	Feature Ad	Discrete(Q_2 = 2)	0 (indifferent), 1 (feature ad on)
$z_{h3}^{(2)}$, $z_{h3}^{(3)}$	Display	Discrete(Q_3 = 2)	0 (indifferent), 1 (display on)
$z_{h4}^{(2)}$, $z_{h4}^{(3)}$	Net-paid Price	Continuous	[\$2.82, \$10.48]

3.2 Estimation Results

All models listed in Table 0 were estimated by MCMC – see Appendix for technical details. Model M_1 is readily estimated by skipping the MCMC steps for s, π_1, Δ, $z^{(2)}$, $z^{(3)}$, $\Psi^{(2)}$ and $\Psi^{(2)}$ under the restriction s_{ht} = 1 for $\forall h,t$. Similarly, M_2 and M_3 can be estimated by skipping MCMC steps for s, π_1, Δ, β, μ_β and Σ_β under the condition s_{ht} = 2 or 3, for $\forall h,t$, and M_4 by skipping the step for Δ under the restriction $\Delta = I_3$.

After 5,000 iterations, all models seemed to reach convergence, judged by monitoring Geweke's (1992) convergence diagnostic. Across all models, the proportion of quantities that passed Geweke's convergence diagnostic ranged from 94.7% to 100%. All inferences reported here are based on the next 15,000 iterations.

3.2.1 Choice of the Most Preferred Model

After estimating M_1 – M_5, integrated likelihoods were computed. The computed integrated likelihoods and the Bayes factors for comparison of models M_1 and M_i (BF_{M_1,M_i}) are given in Table 4.

The first hypotheses to test is:

$$H_0^{(a)}: \text{Heuristics are homogeneous across households } (M_1, M_2, M_3)$$
$$H_1^{(a)}: \text{Heuristics are heterogeneous across households } (M_4).$$

Because BF_{M_i,M_4}, i = 1, 2, 3, are all smaller than 1, households' heuristics usage is apparently not homogeneous. However, among the three homogeneous models, M_1 – M_3, the linear compensatory

Table 4. Model comparison on training sample

M_i	Log of Integrated Likelihood	Log of Bayes Factor $(\log(BF_{M_1,M_i}))$
M_1	–5041.66	0
M_2	–8226.84	3185.17
M_3	–6789.82	1748.15
M_4	–4509.07	–532.6
M_5	–2976.69	–2064.98

model, M_1, is preferred, suggesting that a randomly chosen household may be most likely to invoke the linear compensatory rule on a typical choice occasion. However, the heterogeneous heuristics model, M_4, is decisively preferred even to M_1. The posterior mean of mixing proportion of π_1 in M_4 were (0.7845, 0.0552, 0.1603), with standard deviations (0.0257, 0.0175, 0.0253), implying that 78.5% of households, ceteris paribus, appear to use the linear compensatory rule. All told, then, the null hypothesis $H_0^{(a)}$ was rejected.

The second hypothesis to test is

$H_0^{(b)}$: Heuristics do not change across purchase occasions (M_4)

$H_1^{(b)}$: Heuristics do change across purchase occasions (M_5).

Since $\log(BF_{M_4,M_5}) = -1532.38$, the proposed dynamic hidden Markov mixture choice model, M_5, is greatly preferred to M_4, implying that households indeed adopt different heuristics across their purchase occasions.

Table 4 offers rather strong evidence supporting both heterogeneity and dynamics in heuristics usage. In the sequel, we shall examine M_5 in more detail. Throughout the discussion, statistical inferences on $\alpha = a$ vs $\alpha \neq a$ for a quantity α will be made by examining whether the [2.5 percentile, 97.5 percentile] interval of the posterior distribution of α includes a. In addition, a statistical meaningful difference between two quantities, α and γ, will be determined by comparing a statistic $\tau(\alpha,\gamma) = \dfrac{m_\alpha - m\gamma}{\sqrt{\upsilon_\alpha + \upsilon_\gamma}}$ against a critical value ±1.96, where m_i and υ_i denote the posterior mean and the posterior variance of quantity i.

3.2.2 Estimates of π_1 and Δ

As given in Table 5, the proportions of households that used

Table 5. Estimate of π_1

	Estimate (std. dev.) [2.5 percentile, 97.5 percentile]
$\Pr(s_{h1} = 1)$	0.7322 (0.0364) [0.6578, 0.8007]
$\Pr(s_{h1} = 2)$	0.0561 (0.0369) [0.0040, 0.1449]
$\Pr(s_{h1} = 3)$	0.2116 (0.0520) [0.1043, 0.3076]

Table 6. Estimate of Δ

$s_{h,t-1}$ \ $s_{h,t}$	1	2	3
1	0.8605*(0.0106)** [0.8425,0.8776]***	0.0119(0.0059) [0.0035,0.0227]	0.1276(0.0113) [0.1093,0.1468]
2	0.4309(0.1608) [0.1659,0.6978]	0.1759(0.1270) [0.0169,0.4288]	0.3932(0.1563) [0.1367,0.6567]
3	0.5579(0.0442) [0.4871,0.6317]	0.0474(0.0258) [0.0120,0.0959]	0.3948(0.0519) [0.3085,0.4770]

Note: *: posterior mean; **: posterior std. dev.; ***: [2.5 percentile, 97.5 percentile] interval

linear compensatory, disjunctive and conjunctive rules on their first purchase occasions were 73.2%, 5.6%, and 21.2%, respectively. In order to test whether there is a statistically meaningful difference in the estimates of π_1 across M_5 and M_4, we computed the quantities $\tau(\hat{\pi}_{1,n}^{(4)}, \hat{\pi}_{1,n}^{(5)})$, where $\hat{\pi}_{1,n}^{(i)}$ denotes the estimate of the n-th element of π_1 under model M_i. The computed quantities were (1.17, –0.02, –0.88), implying that the estimates of π_1 under M_4 and M_5 are essentially the same.

Table 6 gives the estimate of the Markov transition matrix, Δ, showing that there were non-zero probabilities for switching among three heuristics. For example, if a household used the linear compensatory rule on its last purchase occasion, the household is likely to use the linear compensatory, disjunctive and conjunctive rules on the current purchase occasion with probabilities {0.86, 0.01, 0.13}. Across all possible states at the last purchase occasion, $s_{h,t-1}$, transition probabilities to the linear compensatory rule, $p(s_{h,t} = 1 \mid s_{h,t-1})$, are noticeably higher than those to the other non-compensatory rules, $p(s_{h,t} = 2 \mid s_{h,t-1})$ and $p(s_{h,t} = 3 \mid s_{h,t-1})$.

It is also interesting to examine whether the proportions of households for three heuristics are close to an equilibrium. If a large number of consumers in the data set are still learning about adequate choice heuristics, the proportions are unlikely to have reached an equilibrium. However, since our data set can be characterized by mundane purchases and there was not a statistically meaningful difference in the estimates of π_1 across M_4 and M_5, the proportions are expected have approached to an equilibrium. To examine this, we computed $\pi_\infty = \lim_{t\to\infty}\pi_t$, where $\pi_t = \Delta\pi_{t-1}$, given MCMC chains of π_1 and Δ (cf., Amemiya 1986). The estimate of π_∞, $\hat{\pi}_\infty$, was (0.7955, 0.0221, 0.1824) with standard deviation (0.0118, 0.0071, 0.0134). Finally, the values of $\tau(\hat{\pi}_{\infty,n}, \hat{\pi}^{(5)}_{\infty,n})$ were (1.65, –0.90, –0.54), implying that π_1 and π_∞ do not differ.

3.2.3 Estimates of s

The proposed hidden Markov dynamic mixture choice model, M_5, allows one to estimate s_{ht} for each $h = 1, \ldots, H$ and $t = 1, \ldots, T_h$. Figure 1 is the histogram of simulated values of $s_{492,2}$ and $s_{492,6}$ and serves as a representative example. The modal values of $s_{492,2}$ and $s_{492,6}$ were 3 and 1, respectively, suggesting that household $h = 492$ likely used conjunctive and linear compensatory rules on its second and sixth purchase occasions, respectively. Figure 2 presents plots of estimates of $p(s_{ht} = i)$, $i = 1, 2, 3$, for four randomly chosen households. All households (with the exception of $h = 137$) displayed varying heuristics usage across their purchase occasions.

Next, we examined the modal values of s_{ht}, $\hat{s}_{ht}$, for each $h = 1, \ldots, H$ and $t = 1, \ldots, T_h$. Figure 3 presents a plot of $\{\hat{s}_{ht}\}_{t=1}^{T_h}$ for $h = 189$ and 215, which indicates these two households did use different choice heuristics across their purchase occasions. In particular,

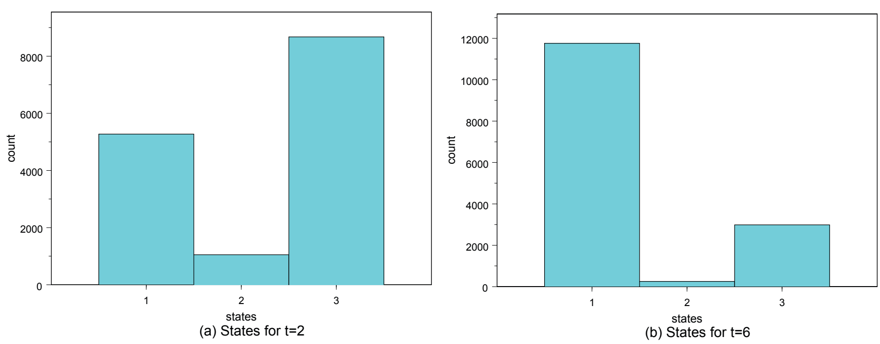

Figure 1. Histogram of simulated values of $s_{492,2}$ and $s_{492,6}$

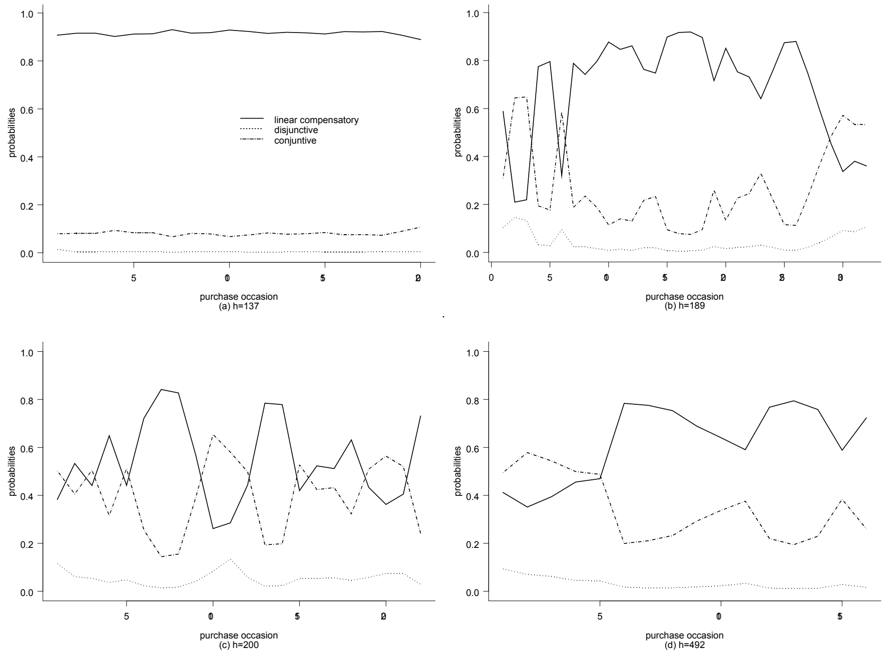

Figure 2. Plot of posterior probabilities of states across purchase occasions

they tended to use the linear compensatory rule on most purchase occasions, but frequently switched to the conjunctive rule. Analogous examination conducted for other households indicates that, among 492 households, 259 households (52.6%) showed changes in $\hat{s}_{ht}$ across purchase occasions. The remaining 233 households used only the linear compensatory rule. Examination of $\hat{s}_{ht}$ thus supports the dynamics of heuristics for a large number of households, and further suggest that a sizeable proportion's choice patterns are consistent with the linear compensatory rule exclusively.

3.2.4 Estimates of Parameters Associated with Three Choice Heuristics

Estimates of Regression Coefficients of the Linear Compensatory Rule

First, let us examine the parameters associated with the linear

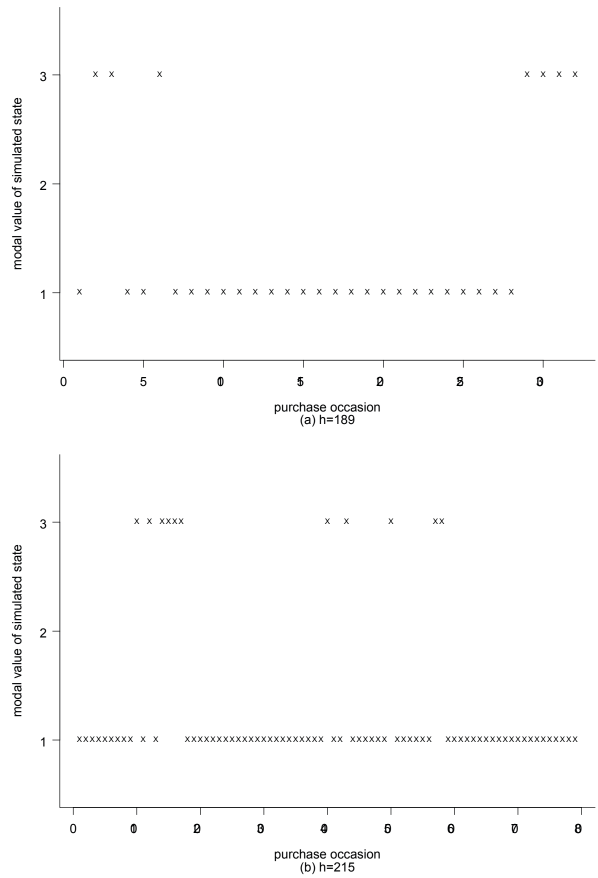

Figure 3. Changes in states across purchase occasions

compensatory rule. Table 7 lists estimates of μ_β and the diagonal elements of Σ_β. All elements of μ_β except the coefficient of the brand C dummy variable were significantly different from zero (i.e., [2.5 percentile, 97.5 percentile] intervals do not include zero). All estimates of feature ad, display and price showed expected signs.

In order to invoke the linear compensatory rule, M_1, a household would need to conduct careful comparisons of all attributes across all brands. However, if a subset of households used other non-compensatory heuristics, M_1 is likely to under- or over-estimate the magnitude of true regression coefficients. Figure 4 is a hypothetical illustration of this point. Suppose that there is only one continuous attribute in utility functions and that households

Table 7. Estimates of μ_β and diag(Σ_β)

	μ_β	diag(Σ_β)
Brand A Dummy	-1.1208* (0.2263)** [-1.5673,-0.6829]***	12.5952(1.4166) [10.0400,15.5909]
Brand B Dummy	1.4596(0.2556) [0.9558,1.9550]	16.3709(1.7465) [13.2495,20.0509]
Brand C Dummy	0.1511(0.2716) [-0.3932,0.6758]	18.6030(1.9188) [15.1018,22.6175]
Feature Ad	0.7038(0.2200) [0.2738,1.1371]	8.8266(1.0300) [6.9526,11.0164]
Display	0.8808(0.2242) [0.4421,1.3207]	9.4431(1.1079) [7.4537,11.7667]
Price	-4.1210(0.2253) [-4.5649,-3.6747]	13.4872(1.3656) [10.9905,16.3533]

Note: *: estimate; **: std.dev.; ***: [2.5 percentile, 97.5 percentile] interval

are divided into two groups in terms of heuristics usage: group C uses the linear compensatory rule, while group NC uses the non-compensatory rules. The lines U_C and U_{NC} depict the hypothetical true utility functions for groups C and NC, respectively. Note that the utility function of the group NC is a step function, given a cut-off w. In this case, if a researcher fits M_1 to an aggregate data set containing both groups of consumers, the estimated utility function may underestimate the true utility function U_C – see U^*. Note that intercepts are same for both U_C and U^*, as implied by M_1. It is similarly straightforward to lay out a case where regression coefficients are over-estimated by M_1.

To examine this possibility, the estimate of μ_β under M_5 was compared with that under M_1 (note that the estimate of μ_β under M_5 is obtained only from households that used the linear compensatory rule). As in Table 8, which lists the relevant comparison figures, the absolute values of all estimates of μ_β under M_1 was smaller than those under M_5, implying that the effects of marketing activities under M_1 might systemically be underestimated. In order to further examine whether the differences were statistically meaningful, values of $\tau(\beta_n^{(1)}, \beta_n^{(5)})$ are listed in the table, where $\beta_n^{(i)}$ is the estimate of the n-th element in β under model M_i. The values of $\tau(\beta_n^{(1)}, \beta_n^{(5)})$ suggest that there was a statistically meaningful difference in the estimates of the price coefficients across the two models. That is, M_1 might well

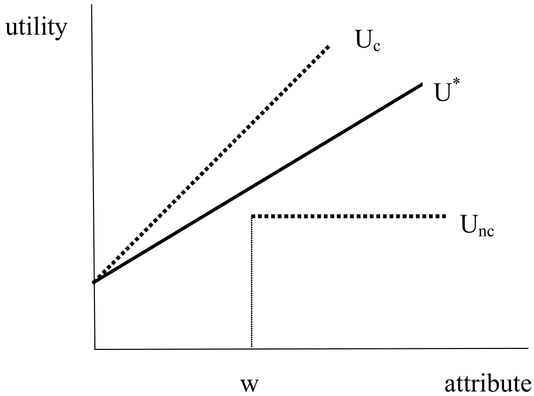

Figure 4. An example of the under-estimation case for regression coefficients in M_1

Table 8. Comparison of estimates of μ_β between M_1 and M_5

	Under M_1	Under M_5	$\tau(\beta_n^{(1)}, \beta_n^{(5)})$
Brand A dummy	–0.8989(0.1843)	–1.1208(0.2263)	-.7605
Brand B dummy	1.0738(0.2033)	1.4596(0.2556)	–1.1814
Brand C dummy	–0.0922(0.2228)	0.1511(0.2716)	–0.6926
Feature ad	0.5276(0.1790)	0.7038(0.2200)	–0.6215
Display	0.6602(0.1902)	0.8808(0.2242)	–0.7505
Price	–2.7716(0.1629)	–4.1210(0.2253)	4.8535

under-estimate regression coefficients for price.

Estimates of Cutoff Levels for the Non-compensatory Heuristics

Rather than reporting the summaries of the population distribution of cutoffs, we examine the estimates, $z_h^{(2)}$ and $z_h^{(3)}$, directly in order to investigate the cutoff phenomenon in more detail. To examine the posterior distribution of $z_h^{(2)}$ and $z_h^{(3)}$, simulated values of $z_h^{(2)}$ and $z_h^{(3)}$ were recorded across MCMC iterations. Next, estimates of $\{z_{h,i}^{(2)}\}_{i=1,2,3}$ and $\{z_{h,i}^{(3)}\}_{i=1,2,3}$ for household h, $\{\hat{z}_{h,i}^{(2)}\}_{i=1,2,3}$ and $\{\hat{z}_{h,i}^{(3)}\}_{i=1,2,3}$ were found using modal values from their MCMC chains. Finally, estimates of $z_{h,4}^{(2)}$ and $z_{h,4}^{(3)}$, $\hat{z}_{h,4}^{(2)}$ and $\hat{z}_{h,4}^{(3)}$ were obtained by taking averages across their MCMC chains. Figure 5 depicts the histograms of $\{\hat{z}_{h,i}^{(2)}\}_{i=1,2,3}$, $\{\hat{z}_{h,i}^{(3)}\}_{i=1,2,3}$ across households, as well as the density plots of $\hat{z}_{h,4}^{(2)}$ and $\hat{z}_{h,4}^{(3)}$ across households.

These results indicate that households were more likely to contain singleton sets of preferred brands when they used the disjunctive than the conjunctive rule. This makes sense, on its face: households having a single preferred brand may focus on maximizing one desired attribute, such as price. Note that feature ad, display and price clearly affected households' choices in the disjunctive rule. By contrast, for the conjunctive rule, *all* households were completely indifferent to feature ad and display, while the household-specific sets of preferred brands and cutoff prices are the main drivers of purchase decisions. Note that the proportion estimates for the 15 (= $2^4 - 1$) possible brand sets (composed of four brands) were different from a uniform distribution of 1/15, so that there is meaningful variation in terms of the set of preferred brands across households. It is noticeable that the proportion the full brand sets, $\{A, B, C, D\}$ dominated other brand sets in the conjunctive rule. Again, this is not surprising: households which engage in detailed processing, as the compensatory rule requires, are more likely to consider a wider range of brands, all else equal. However, even with the full set of brands, households can well eliminate a subset thereof by comparing offered prices against their cutoffs.

Bettman (1975) argued that cutoffs can vary across the disjunctive and the conjunctive rules. Figure 5 supports his argument for the set of preferred brands, feature ad and display. However, his argument was *not* supported for price variable. The means and standard deviations of cutoff prices across households were \$6.31 and 0.07 in the disjunctive rule and \$6.92 and 0.49 in the conjunctive rule, implying the means of cutoff prices were essentially the same ($\tau = -1.23$).

Let us examine the implication of price cutoffs in more detail. Any prices below the cutoff levels can not affect households' choice decisions if conjunctive or disjunctive rules were used. Therefore, the knowledge of the households' cutoff values are important inputs for pricing decisions; the existence of price tiers is well-documented, as is the asymmetric switching phenomenon, where households react differently to price promotions for high-price and low-price brands (e.g., Bronnenberg and Wathieu 1996). To illustrate usage of the estimated cutoff prices for households, let us consider the conjunctive rule. Given the estimates of household-specific cut-off prices for the conjunctive rule, it is easy to examine the proportion of households for which their cutoff prices are grater than, or equal

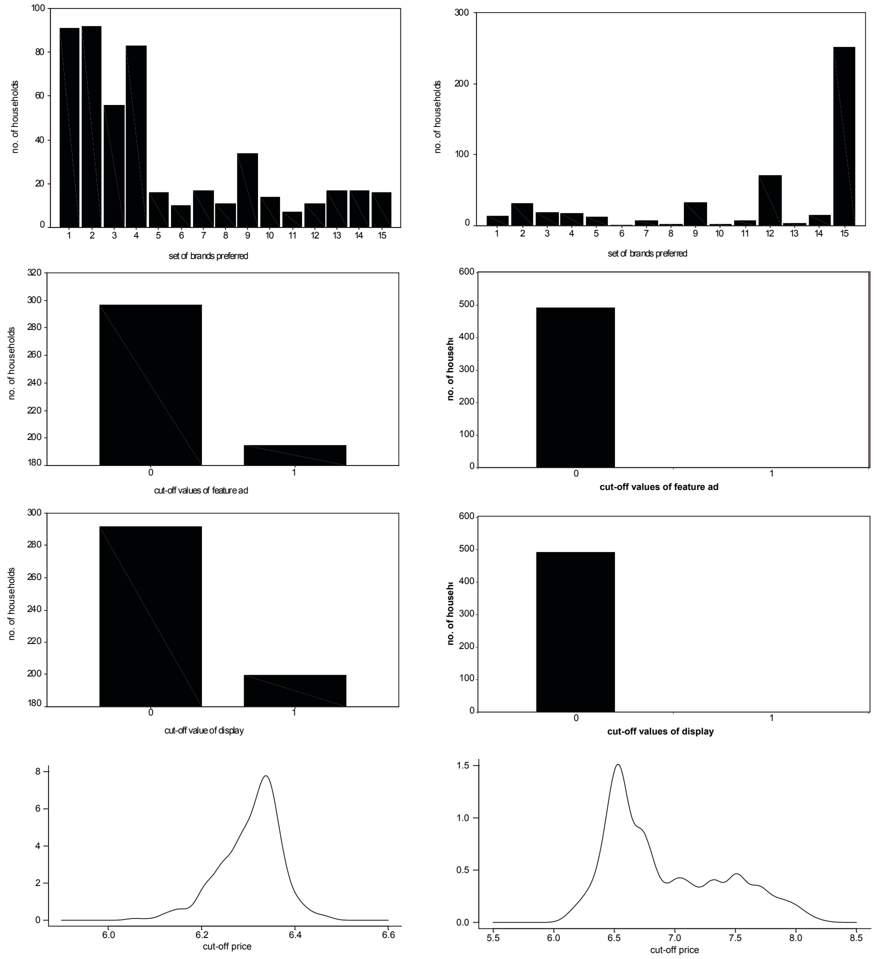

Figure 5. Plot of cut-off values for disjuctive and conjuctive rules

to, a given price ranging from \$2.82 to \$10.48. Figure 6 presents the plot of such proportions. The choice share of a given brand can be increased by increasing the number of households when $E_{ht}^{(3)}$ contains the brand. Figure 13 suggests that all households (among those who use the conjunctive rule) will have the brand in their $E_{ht}^{(3)}$ for prices in [\$2.82, \$5.99]. Certainly, any prices below \$5.99 cannot be optimal.

3.2.5 Causes of Changes in Choice Heuristics over Time

Given the estimate of heuristic variation over time for each

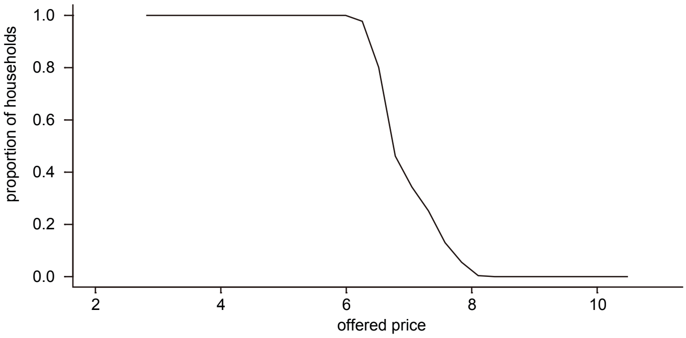

Figure 6. Proportion of households, among those who use the conjunctive rule, when their cut-off prices are greater than, or equal to, offered price

household – the modal values of s_{ht} – one might wonder what drives such changes. In this section, we test a possible explanation for the causes of changes in heuristics over time.

Payne (1982) argued that decision making is contingent upon the characteristics of choice tasks. Specifically, he suggested two types of effects: task effects and context effects. Since structural characteristics of choice tasks, such as the number of alternatives, the number of attributes, time pressure and information display, are fixed or unknown in our data set, we focus instead on possible context effects.

Context effects refers to the influence of general characteristics of decision problems, and reflects the particular values of the alternatives in specific decision sets. Examples of context variables are the similarity of alternatives, the quality of alternative sets and reference points (cf., Payne *et al.* 1993). The notion of similarity has a rich history (Luce 1977, Tversky 1972) and has been hypothesized to affect the information-processing strategies for choice tasks. For example, similarity may affect the degree of ease of comparison among choice alternatives. In this section, we examine the effect of similarity on the usage of heuristics.

We first need to operationalize the notion of similarity. It is obvious that alternatives become more similar, (1) as the variance of attribute values across alternatives decreases, and (2) when the attribute levels have stronger positive correlations. These are among the explicit dimensions used by Shugan (1980) to formalize thinking costs. Since, in our data set, z_h contains categorical variables, the correlation measure is not well defined (although various categorical alternatives do exist). We therefore operationalized similarity as the variance of attribute values across alternatives.

Regarding the effect of similarity on heuristics usage, Biggs *et al.* (1985) proposed interesting predictions. They predicted that, as alternatives become similar, the compensatory rules may be preferred to the non-compensatory ones. They argued that if alternatives are significantly dissimilar, non-comepensatory heuristics are likely to minimize cost of thinking, by allowing decision-makers to quickly eliminate dominated alternatives. However, if alternatives are similar, non-compensatory heuristics may require many comparisons to eliminate alternatives, and compensatory rules are more likely to minimize cost of thinking, since these rules allow choices to be made in 'one pass' by combining weights for attributes with attribute values.

However, the reverse may also be possible. Let us suppose that attribute values across alternatives are very close to one another. In this case, even with well-defined weight structure, it may be difficult to use compensatory rules since (1) the difference in terms of marginal contribution of attributes can be minuscule, and (2) differences in overall evaluation scores of alternatives may also become vanishingly small. In this case, in order to choose a utility-maximizing alternative, decision-makers may need to enact very careful trade-off among attributes. Therefore, purportedly optimizing rules, such as the compensatory, may lead to higher costs and lower benefits than non-compensatory ones. Decision-makers would then rightly prefer the conjunctive or disjunctive rules when alternatives are very similar.

Motivated by the above, we test the competing hypotheses:

H_0: As attribute value variance across alternatives decreases, compensatory rules are preferred.

H_1: As attribute value variance across alternatives decreases, non-compensatory rules are preferred.

Statistical inference was carried out as follows. Given the modal values of z_h, $\hat{s}_{h,t}$, two data sets were constructed:

1. Data I (Sample size: 5528): Collect samples of $\hat{s}_{h,t}$ when $\hat{s}_{h,t-1} = 1$, $t = 1, \ldots, T_h$, $h = 1, \ldots, H$, and define

$$y_{ht}^{(I)} = \begin{cases} 1 \text{ if } \hat{s}_{h,t} \neq s_{h,t-1}, \\ 0, \text{otherwise.} \end{cases}$$

2. Data II (Sample size: 662): Collect samples of $\hat{s}_{h,t}$ when $\hat{s}_{h,t-1} \neq 1$, $t = 1, \ldots, T_h$, $h = 1, \ldots, H$, and define

$$y_{ht}^{(II)} = \begin{cases} 1 \text{ if } \hat{s}_{h,t} \neq s_{h,t-1}, \\ 0, \text{otherwise.} \end{cases}$$

For both data sets, three explanatory variables with an intercept term were introduced to capture similarity among brands:

$$w_{ht} = \begin{bmatrix} 1 \\ v_t(\text{Feature Ad}) \\ v_t(\text{Display}) \\ v_t(\text{Price}) \end{bmatrix},$$

where $v_t(\mathrm{C})$ is a sample variance of C across brands at t.

Next, a probit model was fitted to both data sets. The prior distribution for the regression coefficients of w_{ht} was $N_4(0, 20I_4)$. Estimation results appear in Table 9. As expected, two data sets evidenced opposite signs. In data set I, all explanatory variables, variances of feature ad, display, and price, are statistically meaningfully different from zero, implying that households tend to switch from the linear compensatory rule to the non-compensatory rules as the variances of feature ad, display and price across brands become smaller. Data set II shows that the coefficient of the variance of price is the only explanatory variable different from zero, suggesting that as the variance of price becomes larger, households tend to switch from the non-compensatory rules to the linear compensatory rule.

In summary, Table 9 shows that households in our data set prefer compensatory rules to non-compensatory ones when attributes are more varied. When attribute variance is small, non-compensatory rules are likely to be preferred, and H_1 is supported over H_0.

Table 9. Probit result for Data I and II

	Data I	Data II
Intercept	–1.1646* (0.0442)** [–1.2518, –1.0778]***	0.1555 (0.0848) [–0.0100, 0.3232]
Var. of feature ad	–0.87235 (0.1940) [–1.2593, –0.4961]	0.6527 (0.4016) [–0.1342, 1.4401]
Var. of display	–0.4752 (0.2059) [–0.8758, –0.0658]	0.6558 (0.4301) [–0.1853, 1.4936]
Var. of price	–0.2600 (0.0626) [–0.3837, –0.1376]	0.3141 (0.1103) [0.1017, 0.5326]

Note: *: estimate; **: standard deviation; ***: [2.5 percentile, 97.5 percentile] interval

4. CONCLUSIONS AND FUTURE RESEARCH

This study investigated two issues: heterogeneity and dynamics in heuristics usage, using field data. Using a representative set of compensatory and non-compensatory heuristics – the linear compensatory, the disjunctive and the conjunctive rules – a hidden Markov mixture model was developed to capture changes in heuristics for each household over time. A unique feature of this approach is the ability to study consumers in their 'endogenous environment', making purchase decisions in the real world, under the sorts of constraints ubiquitous in such settings.

We can claim two major findings. First, there exists a great deal of heterogeneity in heuristic usage across households. All three heuristics seem to be used, although among them the linear compensatory is by far the most widely applied, followed by the conjunctive and disjunctive rules. Second, and perhaps most important, heuristics are evidently changing over time at the household-level. To our knowledge, this is the first study to demonstrate anything like this result using real choice data.

Several findings of immediate interest to managers emerged as well. Many marketing studies have shown that ignoring certain empirical effects – purchase acceleration, store-switching, choice endogeneity, among others – can systematically bias the measured strength of marketing variables. Similarly, here we found that the effects of marketing activities can be systematically under-estimated if heterogeneity in heuristic usage is not accounted for. We believe this is another form of heterogeneity which should enter modelers' arsenal, given that ignoring it introduces predictable biases.

Perhaps most relevant for managerial practice, estimates of individual-level cutoffs are of clear use in marketing decision-making. A large corpus of research in marketing has considered reference or reservation prices (e.g., Kalyanaram and Winer 1995). Although we have not incorporated them formally here with the generality prior research has envisioned their effects, the cutoff formulations in the two non-compensatory models act for all intents and purposes as reservation levels. An empirical plot of cutoffs for various attributes across households provides managers with estimates of levels which their target audience deems excessive. In this paper, we can supply such levels not only for prices, but any marketing mix variable.

Finally, we found that the 'compensatoriness' of invoked heuristics is related to how much variation there is across attributes. In this way, the proposed model helps formalize measures used in the cost-of-thinking literature, and provides a method for measuring their effects in real choice scenarios. Specifically, echoing a finding going back to Shugan (1980), we found that high attribute variance favors the use of compensatory heuristics. We believe it would be fruitful to see whether this finding holds up in products categories with markedly different characteristics and purchase behaviors. In addition, we have not attempted to systematically account for multiple possible drivers of heuristic usage change over time, but would expect that variations in environment, needs, household stock, usage occasions or even marketing mix could all be relevant correlates. Differentiating among such causes is certainly a fertile area for future investigations.

The heuristics considered in this paper hardly include all possible heuristics. In some sense, the space of possible heuristics is limitless, and it would be a boon to research in the area if some 'basis' for that space could be rationally or empirically proposed. We have taken a tentative first step in that direction with our three reference heuristics. Another avenue for further research would be to examine how the degrees of heterogeneity and heuristic switching change as more heuristics are explicitly modeled. The proposed hidden Markov mixture model can accommodate any finite number of heuristics so long as they are formulated in mathematically precise language. So we would hope that the general scheme of hidden Markov mixture modeling facilitates increased empirical attention on heuristic usage by consumers in the real world.

References

Amemiya, T. (1986), *Advanced Econometrics*, Oxford, UK: Basil Blackwell Ltd.

Andrews, R. L. and A. K. Manrai (1998), "Feature-based Elimination: Model and Empirical Comparison," *European Journal of Operational Research*, 111, 248-267.

Bernardo, J. M. and A. F. M. Smith (1994), *Bayesian Theory*, New York: Jon Wiley and Sons Ltd.

Bettman, J. R. (1979), "An Information Processing Theory of Consumer Choice," Reading, MA: Addison-Wesley Publishing Company.

Biggs, S. F., J. C. Bedard, B. C. Gaber, and T. J. Linsmeier (1985), "The Effects of Task Size and Similarity on the Decision Behavior of bank Loan Officers," *Management Science*, 31, 8, 970-987.

Bronnenberg, B. J. and L. Wathieu (1996), "Asymmetric Promotion Effects and Brand Positioning," *Marketing Science*, 15(4), 379-394.

Chiang, J., S. Chib, and C. Narasimhan (1999), "Markov Chain Monte Carlo and Models of Consideration Set and Parameter Heterogeneity," *Journal of Econometrics*, 89, 223-248.

Gaudry, M. J. I. and M. G. Dagenais (1979), "The Dogit Model," *Transportation Research*, Series B, 13B, 105-111.

Gensch, D. H. (1987), "A Two-Stage Disaggregate Attribute Choice Model," *Marketing Science*, 6, 223-231.

Geweke, J. (1992), "Evaluating the Accuracy of Sampling-Based Approaches to Calculation of Posterior Moments," in J. M. Bernardo, A. P. Dawid, and A F. M. Smith eds. *Bayesian Statistics 4,* Oxford, UK: Oxford University Press, 169-193.

Gilbride, T. J. and G. M. Allenby (2004), "A Choice Model with Conjunctive, Disjunctive, and Compensatory Screening Rules," *Marketing Science*, 23, 391-406.

Gilbride, T. J. and G. M. Allenby (2006), "Estimating Heterogeneous EBA and Economic Screening Rule Choice Models," *Marketing Science*, 25, 494-509.

Kalyanaram, G. and R. S. Winer (1995), "Empirical Generalizations from Reference Price Research," *Marketing Science*, 14 (Part 2), 161-169.

Luce, R. (1977), "The Choice Axiom After Twenty Years," *Journal of Mathematical Psychology*, 15, 215-233.

Manrai, A. K. and R. L. Andrews (1998), “Two-Stage Discrete Choice Models for Scanner Panel data: An Assessment of Process and Assumptions,” *European Journal of Operational Research*, 111, 193-215.

Manrai, A. K. and P. Sinha (1989), “Elimination-By-Cutoffs,” *Marketing Science*, 8, 133-152.

Payne, J. W. (1982), “Contingent Decision Behavior,” *Psychological Bulletin*, 92, 382-402.

Payne, J. W., J. R. Bettman, and E. J. Johnson (1993), *The Adaptive Decision Maker*, New York: Cambridge university Press.

Roberts, J. H. and J. M. Lattin (1991), “Development and Testing of A Model of Consideration Set Composition,” *Journal of Marketing Research*, 28:429–440.

Rubin, D. B. (1987), “A Noniterative Sampling/Importance Resampling Alternative to the Data Augmentation Algorithm for Creating a Few Imputations When Fractions of Missing information Are Modest: the SIR Algorithm,” comment to a paper by Tanner and Wong, *Journal of American Statistical Association*, 82, 543-546.

Rubin, D. B. (1988), “Using the SIR Algorithm to Simulate Posterior Distributions (with discussion),” in *Bayesian Statistics 3* (eds J.M. Bernardo *et al.*), Oxford: Oxford University Press, 395-402.

Shugan, S. M. (1980), “The Cost of Thinking,” *Journal of Consumer Research*, 7, 99-111.

Siddarth, S., R. E. Bucklin, and D. G. Morrison (1995), “Making the Cut: Modeling and Analyzing Choice Set Restriction in Scanner panel data,” *Journal of Marketing Research*, 32, 255-266.

Swait, J. D. and M. Ben-Akiva (1987), “Empirical Test of a Constrained Choice Discrete Model: Mode Choice in Sao Paulo, Brazil” *Transportation Research*, Series B, 21B, 103-115.

Simon, H. H. (1957), *Models of Man*, New York: John Wiley & Sons, Inc.

Tversky, A. (1972), “Elimination by Aspects: A Theory of Choice,” *Psychological Review*, 3, 213-225.

Wright, P. L. (1975), “Consumer Choice Strategies: Simplifying vs. Optimizing,” *Journal of Marketing Research*, 11, 60-67.

APPENDIX: MCMC ESTIMATION

Define the following quantities:

- $y_h = (y_{h1}, \ldots, y_{hT_h}), y = (y_1, \ldots, y_H)'$,
- $s_h = (s_{h,1}, \ldots, s_{hT_h}), s = (s_1, \ldots, s_H)'$,
- $\mathcal{T}_h^i = \{t \mid s_{h,t} = i, t = 1, \ldots, T_h\}, i = 1, 2, 3$,
- $\mathcal{H}_i = \{h \mid s_{ht} = i \text{ for } \exists t = 1, \ldots, T_h\}, i = 1, 2, 3$,
- $\#(A)$ = the size of a set A,
- $z^{(i)} = \{z_h^{(i)} \mid h \in \mathcal{H}_i\}, i = 2, 3.$
- $\beta = (\beta_1', \ldots \beta_h')'.$

The posterior distribution is

$$p(s, \pi_1, \Delta, \beta, z^{(2)}, z^{(3)}, \mu_\beta, \Sigma_\beta, \Psi^{(2)}, \Psi^{(3)} \mid y) \propto \left(\prod_{h=1}^{H} \prod_{t=1}^{T_h} p(y_{ht} \mid s_{ht}, \phi_h^{(i)}) \right) \times \quad (18)$$

$$\left[\prod_{h=1}^{H} \left\{ p(s_{h1} \mid \pi_1) \prod_{t=1}^{T_h} p(s_{ht} \mid s_{h,t-1}, \Delta) \right\} \right] \times$$

$$\left(\prod_{\mathcal{H}_1} p(\beta_h \mid \mu_\beta, \Sigma_\beta) \right) \times$$

$$\left(\prod_{\mathcal{H}_2} p(z_h^{(2)} \mid \Psi^{(2)}) \right) \times$$

$$\left(\prod_{\mathcal{H}_3} p(z_h^{(3)} \mid \Psi^{(3)}) \right) \times$$

$$p(\pi_1) \times p(\Delta) \times p(\mu_\beta) \times p(\Sigma_\beta) \times$$

$$p(\Psi^{(2)}) \times p(\Psi^{(3)}).$$

To implement a MCMC sampler, we must sample from the following conditional posterior distributions in a sequence,

1. Sampling from $p(s \mid \pi_1, \Delta, \beta, z^{(2)}, z^{(3)}, \mu_\beta, \Sigma_\beta, \Psi^{(2)}, \Psi^{(3)}, y)$: For each h and t, $p(s_{ht} \mid s_{h1}, \ldots, s_{h,t-1}, s_{h,t+1}, \ldots, s_{hT_h}, \pi_1, \Delta, \beta, z \mid y) \propto p(y_{ht} \mid s_{ht},$

$\phi_h^{(i)})p(s_{ht} \mid s_{h,t-1}, \pi_1, \Delta)p(s_{h,t+1} \mid s_{ht}, \pi_1, \Delta)$. Sample $s_{ht} \in \{1, 2, 3\}$ with probabilities (r_1, r_2, r_3) where r_i are normalized probabilities of $(\bar{r}_1, \bar{r}_2, \bar{r}_3)$, where

$$\bar{r}_i = \begin{cases} p(y_{ht} = j \mid s_{ht} = i, \phi_h^{(i)})p(s_{ht} = i \mid \pi_1)p(s_{h,t+1} \mid s_{ht} = i, \Delta), t = 1, \\ p(y_{ht} = j \mid s_{ht} = i, \phi_h^{(i)})p(s_{ht} = i \mid s_{h,t-1}, \Delta)p(s_{h,t+1} \mid s_{ht} = i, \Delta), t = 2, \ldots, T_h - 1 \\ p(y_{ht} = j \mid s_{ht} = i, \phi_h^{(i)})p(s_{ht} = i \mid s_{h,t-1}, \Delta), t = T_h. \end{cases}$$

Note that $\phi_h^{(i)}$, i = 1, 2, 3, must be sampled if the state i is not present among $(s_{h1}, \ldots, s_{h,t-1}, s_{h,t+1}, \ldots, s_{hT_h})$. In this case, the values of $\phi_h^{(i)}$ can easily be sampled by the sampling/importance resampling (SIR) method (Rubin 1987, 1988) as follows: (1) sample $u_{hn}^{(i)}$ from a proposal distribution $q(\phi_h^{(i)})$, n = 1, ..., N, and (2) sample one value among $u_{h1}^{(i)}, \ldots u_{hN}^{(i)}$, with probabilities

$$\left(\frac{\pi(u_{h1}^{(i)})/q(u_{h1}^{(i)})}{\sum_{l=1}^{N}(\pi(u_{hl}^{(i)})/q(u_{hl}^{(i)}))}, \ldots, \frac{\pi(u_{hn}^{(i)})/q(u_{hn}^{(i)})}{\sum_{l=1}^{N}(\pi(u_{hl}^{(i)})/q(u_{hl}^{(i)}))} \right) \cdot \text{ where}$$

$$\pi(u_{hn}^{(i)}) = p(y_{ht} = j \mid s_{ht} = i, \phi_h^{(i)})p(u_{hn}^{(i)} \mid \Psi^{(i)}).$$

2. Sampling from $p(\pi_1 \mid s, \Delta, \beta, z^{(2)}, z^{(3)}, \mu_\beta, \Sigma_\beta, \Psi^{(2)}, \Psi^{(3)}, y)$: $p(\pi_1 \mid s, y) \propto \left({}_{h=1}^{H} p(s_{h1} \mid \pi_1) \right) p(\pi_1) = Dir(\bar{l}_1, \bar{l}_2, \bar{l}_3)$, where $\bar{l}_i = l_i + \#(A_i), i = 1, 2, 3$, with $A_i = \{h \mid s_{h1} = i\}$.

3. Sampling from $p(\Delta \mid s, \pi_1, \beta, z^{(2)}, z^{(3)}, \mu_\beta, \Sigma_\beta, \Psi^{(2)}, \Psi^{(3)}, y)$: For n = 1, 2 , 3 , $p(\delta_m \mid s, y) \propto \left({}_{h=1}^{H} {}_{t=1}^{T_h} p(s_{ht} \mid s_{h,t-1}, \Delta) \right) p(\delta_n) = Dir(\bar{d}_{n1}, \bar{d}_{n2}, \bar{d}_{n3})$, where $\bar{d}_{ni} = d_{ni} + \#(B_{ni})$ with $B_{ni} = \{(h, t) \mid s_{h,t-1} = n$ and $s_{ht} = i$, h = 1, ..., H, t = 1, ..., $T_h\}$.

4. Sampling from $p(\beta \mid s, \pi_1, \Delta, z^{(2)}, z^{(3)}, \mu_\beta, \Sigma_\beta, \Psi^{(2)}, \Psi^{(3)}, y)$: For each household $h \in \mathcal{H}_1$, $p(\beta_h \mid s, y) \propto \left({}_{T_h^1} p(y_{ht} = j \mid s_{ht} = 1, \beta_h) \right) p(\beta_h \mid \mu_\beta, \Sigma_\beta)$, which is sampled by the SIR method.

5. Sampling from $p(z^{(2)}, z^{(3)} \mid s, \pi_1, \Delta, \beta, \mu_\beta, \Sigma_\beta, \Psi^{(2)}, \Psi^{(3)}, y)$: For each household $h \in \mathcal{H}_i$ $(i = 2\ 3)$, $p(z_h^{(i)} \mid s, \Psi^{(i)}, y) \propto \Big({}_{T_h^i} p(y_{ht} = j \mid s_{ht} = i,$

$\phi_h^{(i)})\Big)\, p(z_h^{(i)} \mid \Psi^{(i)}), k = 1, \ldots, K_c$, which is sampled by the SIR method.

6. Sampling from $p(\mu_\beta \mid s, \pi_1, \Delta, \beta, z^{(2)}, z^{(3)}, \Sigma_\beta, \Psi^{(2)}, \Psi^{(3)}, y)$: $p(\mu_\beta \mid \beta_h, \Sigma_\beta, y) \propto \Big(\ _{\mathcal{H}_1} p(\beta_h \mid \mu_\beta, \Sigma_\beta)\Big)\, p(\mu_\beta) = N_K(\bar{f}_\beta, \bar{V}_\beta)$, where $\bar{f}_\beta = \bar{V}_\beta(\sum_{\mathcal{H}_1} \Sigma_\beta^{-1}\beta_h + V_\beta^{-1} f_\beta)$ and $\bar{V}_\beta = (\#(\mathcal{H}_1)\Sigma_\beta^{-1} + V_\beta^{-1})^{-1}$.

7. Sampling from $p(\Sigma_\beta \mid s, \pi_1, \Delta, \beta, z^{(2)}, z^{(3)}, \mu_\beta, \Psi^{(2)}, \Psi^{(3)}, y)$: $p(\Sigma_\beta \mid \beta_h, \mu_\beta, y) \propto \Big(\prod_{\mathcal{H}_1} p(\beta_h \mid \mu_\beta, \Sigma_\beta)\Big)\, p(\Sigma_\beta) = IW_K(\bar{\upsilon}_\beta, \bar{S}_\beta)$, where $\bar{\upsilon}_\beta = \upsilon_\beta + \#(\mathcal{H}_1)$ and $\bar{S}_\beta = \bar{S}_\beta + \sum_{\mathcal{H}_1}(\beta_h - \mu_\beta)(\beta_h - \mu_\beta)^{-1}$.

8. Sampling from $p(\Psi^{(2)}, \Psi^{(3)} \mid s, \pi_1, \Delta, \beta, z^{(2)}, z^{(3)}, \mu_\beta, \Sigma_\beta, y)$: Let $\Psi_m^{(i)} \in \Psi^{(i)}$ denote parameters associated with m-th elements of $z_h^{(i)}$, $m = 1, \ldots, M$. Then, $p(\Psi_m^{(i)} \mid z, y) \propto \Big(\ _{\mathcal{H}_i} p(z_{hm}^{(i)} \mid \Psi_m^{(i)})\Big)\, p(\Psi_m^{(i)})$. Then, the conditional posterior distributions of $\Psi_m^{(i)}$ are (a) $Dir(\bar{e}_{m1}, , \ldots, \bar{e}_{mQ_n})$, when $z_{hm}^{(i)}$ is a discrete variable, and $N(\mu_m^{(i)} \mid \bar{f}_m^{(i)}, \bar{\upsilon}_m^{(i)})$ and $IG(\sigma_m^{(i)} \mid \bar{l}_m^{(i)}, \bar{g}_m^{(i)})$, when $z_{hm}^{(i)}$ is a continuous variable, where $\bar{e}_{mq} = e_{mq} + \sum_{\mathcal{H}_i} I_{z_{hm}^{(i)} = r_{mq}}, q = 1, \ldots, Q_m$ and $\bar{f}_m^{(i)} = \bar{\upsilon}_m^{(i)}\left(\sum_{\mathcal{H}_i} \frac{z_{hm}^{(i)}}{\sigma_m^{(i)}} + \frac{f_m}{\upsilon_m}\right), \bar{\upsilon}_m^{(i)} = \sigma_m^{(i)}\upsilon_m \big/ \left\{\#(\mathcal{H}_i)\upsilon_m + \sigma_m^{(i)}\right\}, \bar{l}_m^{(i)} = l_m + \frac{\#(\mathcal{H}_i)}{2}$ and $\bar{g}_m^{(i)} = g_m + \frac{1}{2}\sum_{\mathcal{H}_i}(z_{hm}^{(i)} - \mu_m^{(i)})^2$.

Received April 23
Accepted May 28