95-F-26

A Maximum Likelihood Non-Compensatory Model: You Trade-off or Satisfice ?

by

Yutaka Hamaoka
University of Tokyo

Ichiro Furukawa
Hitotsubashi University

Hotaka Katahira
University of Tokyo

October 1995

A Maximum Likelihood Non-Compensatory Model:

You Trade-off or Satisfice ?

Yutaka Hamaoka

(University of Tokyo)

Ichiro Furukawa

(Hitotsubashi University)

Hotaka Katahira

(University of Tokyo)

October 1995

1. **Introduction**

2. **Model**

3. **Estimation and Some Small Sample Properties**

5. **Pilot Applications**

6. **Discussion and Conclusion**

---

1. Introduction

The proliferation of 'logit' papers has characterized marketing science literature in the last decade and a half. Explicitly or implicitly, an individual consumer's decision was assumed to be compensatory there. By compensatory we mean that a consumer's evaluation of an object is such that a low score with one attribute (e.g. too high a price) may be compensated for by a high score with another (e.g. very easy to use). Although there is reasonable theoretical and empirical evidence that a compensatory model is a good approximation to non-compensatory decisions under various conditions[Dawes and Corrigan(1974); Johnson and Meyer(1984)], there is as strong an argument primarily by consumer researchers that individuals do actually employ non-compensatory rules very frequently[Bettman and Jacoby(1976); Payne and Ragsdale(1978)] and that the process can't be approximated by a compensatory model under some conditions, e.g. negative correlation among attributes[Johnson, Meyer and Ghose(1989)].

Given the state of the art in the discussions of compensatory vs. non-compensatory decisions, the unbalanced popularity of compensatory models in empirical marketing studies is difficult to justify solely on its ability of 'approximation'. At a risk of being too wild, the authors suspect that an easy accessibility of the estimation technologies for a compensatory model like logit drove researchers to apply it without scrutinizing the legitimacy of its use.

The scarcity of the applications of non-compensatory models doesn't imply that non-compensatory processes have been studied very little. Far from that, we have seen a considerable amount of efforts particularly in theoretical developments. Lexicographic, conjunctive and disjunctive models are among the simplest and best known [Einhorn(1970)]. Another class of models are developed to deal with sequential elimination processes, including Tversky's Elimination By Aspects(EBA) model [Tversky(1972)]. A simplest deterministic version is known as a gatekeeper model that is actually a sequencial extention of a conjunctive rule and was proved to account for supermarket buyers' buying decisions better than a linear rule [Montgomery (1975)]. Other interesting elimination models are Pretree model[Tversky and Sattath (1979)] and Hierarchical Elimination Model (HEM) [Hauser(1985)]. Among more recent developments is Elimination By Cutoff (EBC) model [Manrai and Sinha (1989)] that is an extention of EBA to a product map setting.

Unlike the theoretical developments, the statistical estimation technologies for non-compensatory models as accessible and as easy to use as ML logit are surprizingly underdeveloped. Of course, there are some models equipped with estimation methods, like an MLH model by Gensch and Svestka(1985) but they either model a very specific decision strategy, require a complicated estimation algorithm or both. To the authors' knowledge, transportation research is the only field where statistical models of non-compensatory processes are developed and applied relatively widely. There, such models are used to describe and predict the screening of alternatives for the formation of a choice set. A random constraint model by Swait and Ben-Akiva (1987) is an example and shares some inspiration with a conjunctive individual version of the present model.

The primary purposes of this paper are first, to present a simple and flexible model of a non-compensatory choice together with an easy estimation procedure, and secondly to demonstrate experimantally and empirically that our non-compensatory model works better than a compensatory (logit) model in terms of fit and external validity under certain conditions.

The next section presents a set of models. Section 3 discusses the maximum likelihood(ML) estimation procedures of the model and presents some results of Monte Carlo simulations on the small sample properties of our ML estimators. The pilot application results are introduced in Section 4 and summary and conclusion is given in the end.

2. Model

The proposed model is a model of noncompensatory choice including conjunctive and disjunctive ones. It is a model of an individual as well as of a heterogeneous aggregate population. It basically deals with GO-NO type binary data but is good for any type of data that could be transformed into a binary type, i.e. ordered category data. As a starter, a base model of conjunctive choice of an individual using binary data is presented. Extentions from the base model are discussed subsequently.

A Base Model: An Individual Conjunctive Model for Binary Data

A base model is a model of an individual making a conjunctive choice. The outcomes are GO and NO given an object with I dimensional attributes. The conjunctive rule assumes that one is GO with object j if all the attributes of j clear his/her bottomline and is NO otherwise. Let y_{jk} be a binary dependent variable taking 1 if individual k is GO with object j and 0 otherwise, x_{ij} ith attribute of object j and τ_{ik} threshold of individual k with attribute i. The conjunctive(satisficing) rule requires that

$$(1) \qquad y_{jk} = 1 \qquad \text{iff} \qquad x_{ij} > \tau_{ik} \quad \text{for all } i$$

$$= 0 \qquad \text{otherwise.}$$

For the time being, x_{ij} is assumed to be positive, continuous and positively related to the dependent variable. Also subscript k is omitted for simplicity in discribing the individual version. Now τ_{ik} is assumed to be a random variable whose randomness comes from intrinsic temporal variation in the individual's threshold level. When the model is aggregated over individuals, we might naturally be able to incorporate heterogeneity in the mean level of thresholds into the randomness of τ_{ik}. This aggregation issue is elaborated on later.

The randomness of τ_i renders y_j random. The probability of $y_j = 1$ is now a joint probability of $x_{ij} - \tau_i > 0$ for all i;

$$(2) \qquad \text{Prob.}(y_j = 1) = \Pi_i \text{ Prob.}(x_{ij} - \tau_i > 0).$$

Assuming that τ_i is i.i.d. logistically distributed with mean t , the probability of x_{ij} exceeding τ_i is given as;

(3) $\quad$ $\text{Prob.}(x_{ij} > \tau_i) = P_{ij} = 1 / \{1 + \exp\text{-}\, \beta_i(x_{ij} - t_i)\}$.

Inserting this into (2) gives

(4) $\quad$ $\text{Prob.}(y_j = 1) = \Pi_i P_{ij} = \Pi_i 1 / \{1 + \exp\text{-}\, \beta_i(x_{ij} - t_i)\} = P_j$.

While t_i is the mean threshold for i th attribute, β_i is inversely proportional to the size of the variance.

To give a deterministic flavor to our construct of stochastic threshold, we might define "95 % threshold", $t_{.95i}$, as follows;

(5) $\quad$ $.95 = 1 / \{1 + \exp\text{-}\, \beta_i(t_{.95i} - t_i)\}$.

Obviously, $t_{.95i}$ is the minimum level of attribute i for satisfaction 95 % of the time. The concept of 95 % threshold is graphically illustrated in Exhibit 1.

(Exhibit 1)

So far the attributes are assumed to be continuous. Although the concept of 'threshold' is only relevant to the continuous attributes, one might like to include discrete variables to explain satisfaction. Let z_{gj} be a dummy variable taking 1 if object j has a discrete attribute g and 0 otherwise. The probability that object j is satisfactory with respect to attribute g , P_{gj} is specified as,

(6) $\quad$ $P_{gj} = \{1 / (1 + \exp\text{-}\, \beta_i)\}\, z_{ij} + \{1 - 1 / (1 + \exp\text{-}\, \beta_i)\}(1 - z_{ij})$

Including these terms for discrete attributes, P_j in eq.(4) are rewritten as,

(7) $$P_j = (\Pi_i P_{ij})(\Pi_g P_{gj}).$$

For the brevity of exposition, the remaining discussion only deals with the case where all the attributes are continuous.

Aggregating Heterogeneous Population

One of the beauties of the present model is its excellent aggregation properties. Unlike the linear logit formulation that requires various complicated and often ad hoc procedures to cope with heterogeneity in parameters including generalized probit [Hausman and Wise(1978)] and cluster-wise logit [Kamakura and Russell(1989)], it is structurally straightforward with the present formulation.

Assuming that the mean of an individual is distributed logistically around the grand mean, τ_{ik} is decomposed as;

(8) $$\tau_{ik} = t_{0i} + \gamma_i + \varepsilon_{ik}$$

γ_i and ε_{ik} being independently logistically distributed with mean 0, the unconditional probability that object j is GO is given by eq.(4) with t_i replaced by t_{0i}.

(9) $$\text{Prob.}(y_j = 1) = \Pi_i \ 1 / [1 + \exp\text{-} \beta_i(x_{ij} - t_{0i})] .$$

Now the small β_i implies either that the intrinsic variation in an individual's threshold is large or that the population is highly heterogeneous with respect to a mean threshold.

Explosion of Ordered Category / Continuous Rating Data

The original data should not necessarily be binary. Rating and other types of data are also good to the extent that they are meaningfully transformed into a binary one. The rating data might be converted to binary by dichotomizing the categories like 'top two box or not'. Let y_{sj} be a transformed dependent variable taking 1 if the object j is rated in the top s categories and 0 otherwise. Also let the corresponding parameters be β_{si} and t_{si}. τ_{si} represents that level of threshold for attribute i which determines whether the object is rated in the top s categories. An interesting feature of this explosion of rating data is illustrated in Exhibit 2. By comparing the levels of, say, $t_{.95si}$ over s for all i, one could find which attribute is critical in order to obtain s th rather than s+1 th rating. The Exhibit tells, for example, that attribute 1 is critical in having 'definitely will buy' response, while attribute 2 is important only to avoid the bottom box. The former might be called 'an enhancing attribute' while the latter is often called 'a screening attribute'.

(Exhibit 2)

A similar explosion is possible, of course, with continuous rating data. 'The s th rating category' is now replaced by 'over s point rating' and the same argument applies.

With either type of explosion, managerially of great interest is the demonstration that the critical attribute is conditional on the level of performance (e.g. the very top vs. better than average) one aspires to achieve. In the case of Exhibit 2, the implied recommendation would be to make sure that attribute 1 be just above x and to spend all the remaining resources on enhancing the level of attribute 2. Unfortunately, a compensatory model is not capable of doing this.

Extention to a Disjunctive Model

The base model is easily extended to accomodate a disjunctive choice. According to the disjunctive rule, one is GO with an object if at least one of the attributes is above its threshold. More specifically, the probability that object j is GO is given as follows;

(10) $$\text{Prob.}(y_j = 1) = 1 - \Pi_i\ [1 - 1 / \{1 + \exp\text{-}\ \beta_i(x_{ij} - t_i)\}].$$

It might be noted that this is structurally very similar to a conjunctive model.

3. Estimation and Some Small Sample Properties

The maximum likelihood estimation procedures for a base model pooling across individuals are introduced and the small sample properties of the estimators are examined via Monte Carlo simulation.

Maximum Likelihood Estimation

Given the data $\{\ y_{jk}\ ; j = 1,\ldots,J,\ k = 1,\ldots,K\ \}$, the likelihood of observing them is;

(11) $$L = \Pi_k \Pi_j P_{jk}^{y_{jk}}\ [1 - P_{jk}]^{1-y_{jk}}$$

where P_{jk} for each k is given by eq.(9). The ML estimates of β_i and $\Delta_i = \tau_i \times \beta_i$ for $i = 1,\ldots,I$ are obtained by maximizing the above likelihood by, say, such standard method as Newton-Raphson algorithm. The estimates of τ_i's are obtained by dividing the estimates of Δ_i's by that of β_i's. In formal terms, this is nothing but a binary logit and its existence and uniqueness conditions equally apply here as well.

Some Monte Carlo Results

From applications point of view, it is of vital importance to examine to what extent a large sample theory can be applied in making inference.

To examine the small sample properties of the ML estimators, a Monte Carlo experiment was designed as follows.

(1) Specify a set of the true values of parameters, β_i*'s and t_i*'s. Emulating the environments of the pilot application of next section in terms of the size of parameters and variables, β_i*'s are uniformly set across i at .01, .05 and .1 and t_i*'s at 50, a mid point for 100 point scale.

(2) Generate x_{ijk} by independently generating uniform random numbers in [0, 100]. Note that this corresponds to a case of 'orthogonal environment' in Johnson, Meyer and Ghose(1989)'s terminology.

(3) Generate Logistic random numbers for ε_{ijk} by generating uniform random numbers in [0,1] and transforming it.

(4) Simulate a choice by comparing x_{ijk} with $\tau_i = t_i^* + \varepsilon_{ik}$ over i and determining y_{jk}

(5) Pooling 1280 observations.

(6) Estimate the parameters using $N = 20, 40, 80, 160, 320, 640$, and 1280 observations out of the pooled 1280, for each set of prespecified parameters. Repeat this 100 times.

(7) Calculate following indicators to evaluate the small sample properties for each case. Let a typical parameter be denoted by θ, its g th estimate by θ_g and the number of repetitions by G. The indicators are,

$$\underline{\text{bias rate}} = \{E(\theta\text{^}) - \theta^*\} / \theta^* = \{(\Sigma_g \theta_g / G) - \theta^*\} / \theta^*$$

$$\underline{\text{t value}} = \{(\Sigma_g \theta_g / G) - \theta^*\} / s(\theta\text{^}), \text{ where}$$

$$s2(\theta\text{^}) = \Sigma_g(\theta_g - \text{Ave.}(\theta))2 / (G - 1)$$

$$\underline{\text{Mean Absolute Error}} = \Sigma_g | \theta_g - \theta^* | / G$$

The results of the experiments are summarized in Exhibit 3.

(Exhibit 3)

Found from the Exhibit were the following.

1 The small sample properties are moderate with $N > 160$ and good with $N > 320$ in terms of bias and MAE.

2 With the sample size fixed, the larger the number of attributes, the more likely is the case that the ML estimators don't exist and the estimates, if existent, get less accurate.

3 With the sample size and the true mean thresholds, τ_i*'s given, the larger the variance of a threshold (the smaller the β_i*'s), the better behaved the estimates are.

Modifying McFadden's Lemma [Lemma 3 and Axiom 6 in McFadden(1974)], a necessary condition for the existence of the ML estimators is that there doesn't exist any solution $z = \{v_i\}$ satisfying the following system of inequalities;

(12) $y_{jk}(x_{ijk} - v_i) + (1 - y_{jk})\,(x_{ijk} - v_i) > 0$; for all j and k

The first part of the point 2 above is exactly related to this. Given the same number of observations(J times K), the solution space for inequlities (12) is more 'open', the larger the number of attributes, I. Consequently, nonexistence of ML estimators is more likely due to this consistency of response data.

To sum, with up to 4 attributes, a smple size of 160 appears an absolute bottomline in terms of well-behaved small sample properties. If one encounters a case of nonexistence of estimators, it is very likely that inequalities (12) hold and the multiple solution space might be obtained by directly solving the system of inequalities.

5. A Pilot Application

To illustrate how the model is estimated using actual data, an application was made to a case of individuals' response to TV commercials. Specifically, the dependent variable was whether the response was among the top two categories in five point advertising attitude scale. The explanatory variables were 100 point ratings of the commercials in terms of various emotional response variables.

The respondents were 48 people(including 1 woman) working at the leading Japanese firms whose average age was approximately 35. They were first shown a series of 8 different commercials, 4 in PC and 4 in beer categories. Then, they were shown each one of them twice and asked to fill in the questionaire, repeating it 8 times for all the commercials. Such criteria as brand recall, brand attitude, purchase intention and others were measured as well as advertising attitude. Twelve variables were used to measure the emotional response to the commercials, including 'upscale image', 'irritating', 'fresh image' etc.

Exhibit 4 illustrates two sets of average ratings in these variables for the four beer commercials; one for those who gave the top 2 box ratings in ad attitude and another for the rest of the sample. From the Monte Carlo results, it is reasonable to use a maximum of 4 variables given the sample size of 192(48 people times 4 commercials). The Exhibit demonstrates that 'amusing', 'suitable cast', 'irritating' and 'familiar' are the four variables that gave rise to the highest differences.

(Exhibit 4)

Estimation Results

Starting with the four explanatory variables, the variables are eliminated one by one by sequentially applying a likelihood ratio test. The test results as well as the final estimation results are shown in Exhibit 5.

(Exhibit 5)

It was found that two variables, 'amusing' and 'familiar', are sufficient to account for the advertising attitude ratings.

Using 'top box', 'top 2 box', etc. for dichotomizing the responses, that is, exploding the rating data into multiple binary response data, the response profiles are obtained for each of the two explanatory variables and shown in Exhibit 6.

(Exhibit 6)

The Exhibit demonstrates that the profile for 'familiar' is linear, while that for 'amusing' is more kinked, implying that a very high score in 'amusing' is required to secure a top box response in advertising attitude. The probability response curves, $P_{ij}(x_{ij})$'s, given in Exhibit 7 show that the distributions of the thresholds for the two variables are quite tight implying that neither inter- nor intra-individual variations were large.

(Exhibit 7)

It might be noted that the use of a logistic distribution to a bounded dependent variable is only an approximation and that extrapolational inference is dangerous.

Comparison with a Linear Model: Holdout Sample Tests

To validate our conjunctive specification and to compare it against a linear compensatory, holdout tests were made estimating both specifications on both test and holdout samples. The test sample was the responces to all but a holdout commercial and the holdout sample was those to the holdout commercial. Kirin Ichiban commercial, the most popular in ad attitude, was used as holdout commercials.

The test results are given in Exhibit 8. Although the fits of our model are slightly worse in terms of log likelihood, the holdout predictions are better in all the cases. The criterion was the observed % of the top box responses as against the average of the estimates of P_{iJk}'s over k, J being the holdout commercial. These results clearly demonstrates that the conjunctive specification gives consistently robust predictions, where a linear compensatory model suffers from greater overestimation of success probabilities.

(Exhibit 8)

6. Summary and Conclusion

A simple model of noncompensatory choice was set forth and an easy method of its estimation by maximum likelihood was proposed. The ML estimators are found to be reasonably well-behaved in a problem of moderate size; 4 attributes and $N > 160$. A pilot study demonstrated that our model explains advertising attitude ratings just as well as a linear binary logit analysis and did clearly better in holdout validation. The model is neither grand nor novel but it is a mere variation of familiar binary logit. It, however, seems to do a number of nice things that a linear compensatory analysis failed to cover.

A contribution of the present study is simple and straightforward. It opened up a path to simple and handy estimation of conjunctive and disjunctive models of consumer choice. Given the 'algorithm dominates theory' phenomena in marketing and consumer studies represented by the linear logit sovereignity, the presentation of our algorithm is hoped to encourage the proliferation of empirical applications and testing of non-compensatory models.

Secondly, the present study empirically demonstrated that our non-compensatory model could be a better 'misspecification' of a real decision process. The finding that our model outperformed a binary logit in predicting ad attitude ratings is particularly promising.

Thirdly, it makes better anatomy of the architecture of satisfaction. Taxonomy of the relevant attributes into various categories including linear, screening and enhancing attributes coupled with the estimation of the structure as such enables to better understand how people are dis/satisfied. This at the same time gives strong policy implications regarding what attributes to improve under specific conditions and objectives.

This is virtually the first easy step to an empirical world of non-compensatory choice. One might easily think of following directions for further research. An obvious direction would be to replicate applications to make the technology a more reliable one. Of particular interest is the further examination of external validity as compared with linear compensatory specifications. In a due course of increasing applications, we will hopefully have better understanding of under what conditions a non-compensatory model describes and predicts better than a compensatory counterpart.

The present model typically takes care of the 'pick any / N' situation. It is therefore very appropriate to use it for explaining the choice set formation processes as Swait and Ben-Akiva(1987) did in transportation research. A natural extention is to integrate the present model with 'pick one / "any"' type of models that explain the choice of an object from the choice set. The idea of two stage model was already empirically explored by Gensch(1987) using his MLH model for the first stage and logit for the second. Very recently, Morikawa(1995) proposed a model of integrating conjunctive choice set formation with logit choice together with a simultaneous estimation algorithm. This is a right direction to follow. Instead, it might also be of interest to apply our conjunctive model directly to purchase records of store scanner panel data. Suppose an individual chose brand A out of *N* different brands of a category on the shelf. We might treat this brand choice as a series of *N* binary decisions of buy or not buy. This offers an additional advantage of being able to take care of multiple choices in a category on the same basis.

Among the remaining problems are elaboration of model structure and/or estimation technologies, development of a powerful method of comparing alternative models (e.g. 2 stage vs. 1 stage) and integration of protocol, experimental, econometric and other research outcomes in the field.

Finally, the authors hope that the present paper shed a small light on the 'dark continent of the choice model' and stimulate a continuing flow of empirical studies in the field.

[REFERENCES]

Bettman, James R. and Jacob Jacoby (1976) , "Patterns of Processing in Consumer Information Acquisition," in *Advances in Consumer Research*, Vol.3, Beverlee B. Anderson, ed. Ann Arbor, MI: Association for Consumer Research, 315-20.

Dawes, Robin M. and Bernard Corrigan (1974), "Linear Models in Decision Making," *Psychological Bulletin*, 81 (March), 95-106.

Einhorn, Hillel J. (1970), "The Use of Nonlinear, Noncompensatory Models in Decision Making," *Psychological Bulletin*, 73 (3), 221-30.

Gensch, D. H. and J. A. Svestka (1984), "A Maximum Likelihood Hierarchical Disaggregate Model for Predicting Choices of Individuals," *Journal of Mathematical Psychology*, 28, 160-178.

Hauser, J. R. (1985), "Agendas and Consumer Choice," working paper No. 1641-85, Alfred P. Sloan School of Management, MIT.

Hausman, J. A. and D. A. Wise (1978), "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences," *Econometrica*, 46, 403-426.

Johnson, Eric J. and Robert J. Meyer (1984), "Compensatory Models of Non-Compensatory Choice Processes: The Effect of Varying Context," *Journal of Consumer Research*, 11 (June), 528-41.

Johnson, Eric J. and Robert J. Meyer, and Sanjoy Ghose (1989), "When Choice Models Fail: Compensatory Models in Negatively Correlated Environments," *Journal of Marketing Research*, Vol.XXVI (August), 255-70.

Kamakura, W. A. and G. J. Russell (1989), "A Probabilistic Choice Model for Market Segmentation and Clasticity Structure," *Journal of Marketing Research*, Vol.26 (November), 379-390.

Manrai, A. K. and P. Sinha (1989), "Elimination-By-Cutoffs," *Marketing Science*, Vol.8, No.2 (Spring), 133-152.

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

Montgomery, D. B. (1975), "New Product Distribution: An Analysis of Supermarket Buyer Decisions," working paper No.205, Graduate School of Business, Stanford University.

Morikawa, Taka (1995), "A Hybrid Probabilistic Choice Set Model with Compensatory and Noncompensatory Choice Rules," paper presented at the 7th World Conference on Transport Research, Sydney.

Payne, John W. and E. K. Easton Ragsdale (1978), "Verbal Protocols and Direct Observation of Supermarket Shopping Behavior: Some Findings and a Discussion of Methods," in *Advances in Consumer Research*, Vol.5, H. Keith Hunt, ed. Chicago: Association for Consumer Research, 571-7.

Swait, Joffre and Moshe Ben-Akiva (1987), "Incorporating Random Constraints in Discrete Models of Choice Set Generation," *Transportation Research Bulletin* Vol. 21B, No. 2, 91-102.

Tversky, Amos (1972), "Elimination by Aspects: A Theory of Choice," *Psychological Review*, 79 (July), 281-99.

Tversky, Amos and Shumel Sattath (1979), "Preference Trees," *Psychological Review*, 86 (November), 542-73.

Exhibit 1 95% Threshold

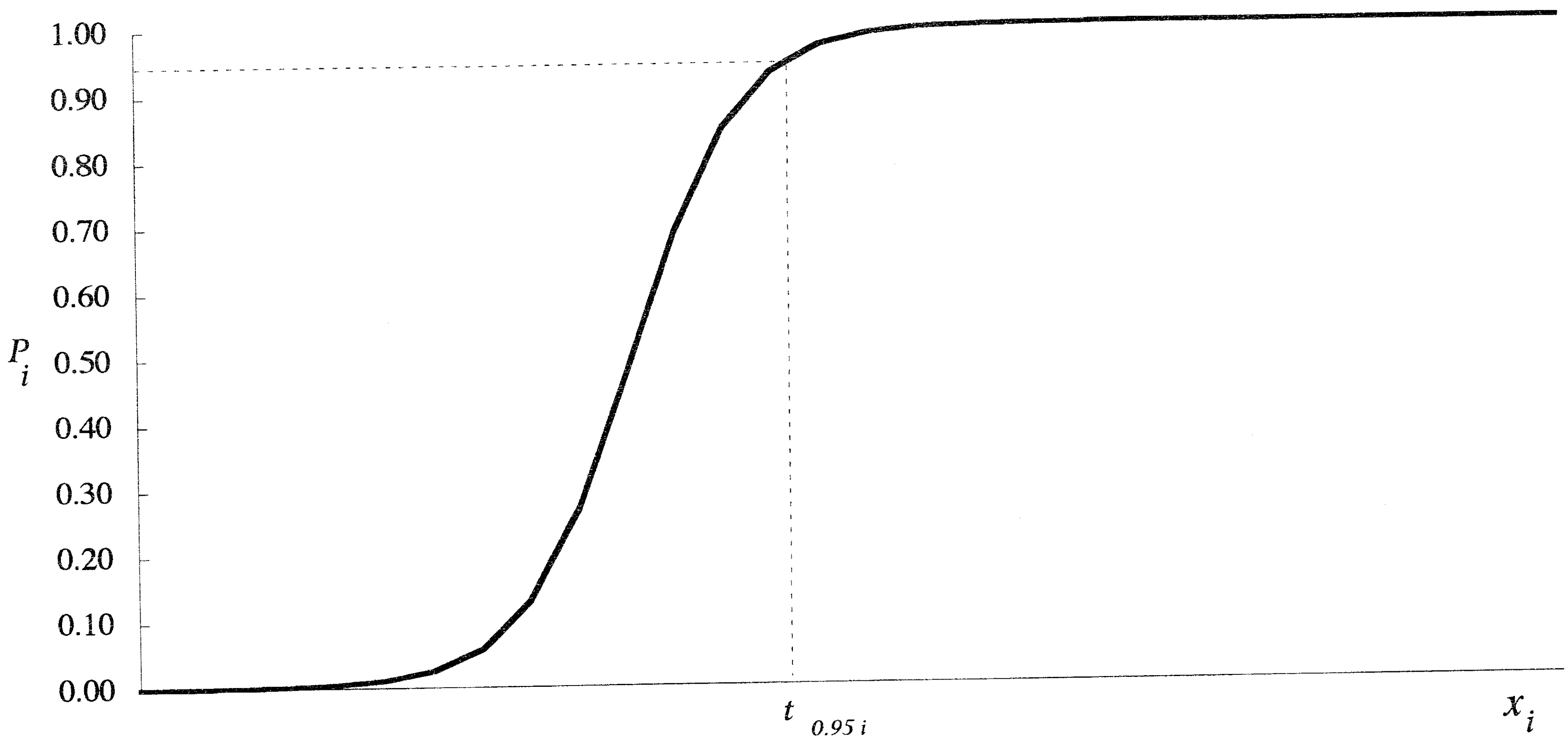


Exhibit 2 Enhancing Attribute and Screening Attribute

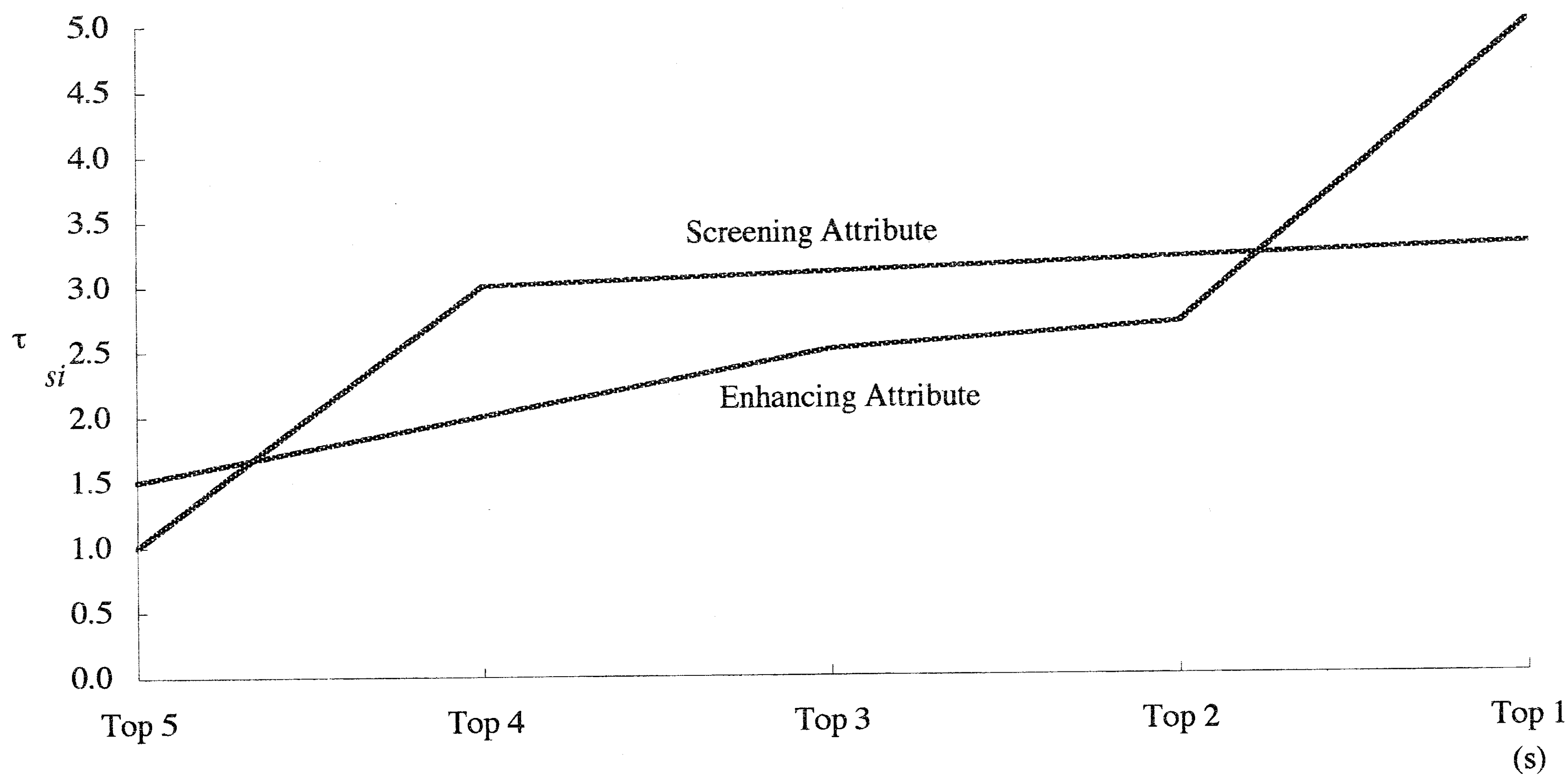


Exhibit 3 Results of Monte Carlo Simulation

(1) βi=0.01, Δi=0.5 (i=1~4, $\tau i=\Delta i/\beta i$=50)

Test Statistics	Number of Obs.	a) 2 attributes				b) 4 attributes							
		$\beta 2$	$\Delta 2$	$\beta 1$	$\Delta 1$	$\beta 4$	$\Delta 4$	$\beta 3$	$\Delta 3$	$\beta 2$	$\Delta 2$	$\beta 1$	$\Delta 1$
Average of Estimated Parameters	20	0.030	-0.312	0.088	3.488	0.047	3.325	-0.019	1.804	0.017	-0.004	0.050	-0.491
	40	0.027	0.803	0.053	1.170	0.040	-1.012	0.010	2.096	-0.007	1.795	0.003	1.725
	80	0.021	0.645	0.017	-0.055	0.035	1.886	0.083	0.399	0.041	0.783	0.025	-0.444
	160	0.024	0.224	0.025	0.405	0.043	0.854	-0.018	-2.593	0.021	0.291	0.059	1.953
	320	0.020	0.383	0.023	0.309	0.008	-0.237	0.022	0.058	0.022	-0.268	0.042	0.120
	640	0.022	0.429	0.013	0.555	0.030	0.449	0.038	-0.112	0.029	-0.377	0.034	0.172
	800	0.020	0.429	0.017	0.587	0.027	0.336	0.031	0.056	0.034	0.638	0.035	0.301
	1280	0.019	0.437	0.020	0.574	0.019	0.409	0.034	0.350	0.034	0.416	0.042	0.605
Standard Error	20	0.080	5.221	0.256	11.486	0.022	1.209	0.051	3.959	0.052	7.524	0.093	8.493
	40	0.056	2.600	0.190	7.737	0.071	8.211	0.083	2.340	0.078	2.931	0.074	2.990
	80	0.041	1.783	0.042	2.652	0.072	4.565	0.232	11.194	0.070	6.303	0.042	6.388
	160	0.070	3.740	0.046	1.657	0.136	3.182	0.119	7.175	0.048	1.979	0.148	5.040
	320	0.049	3.023	0.044	1.292	0.063	4.536	0.050	3.251	0.065	3.040	0.084	3.572
	640	0.027	0.748	0.016	0.774	0.105	3.188	0.093	3.661	0.111	3.673	0.079	3.418
	800	0.023	0.672	0.029	0.627	0.063	1.826	0.085	3.148	0.064	1.338	0.076	3.266
	1280	0.024	0.588	0.037	0.556	0.030	1.577	0.065	2.141	0.056	1.466	0.092	1.174
t-Value	20	0.380	-0.060	0.346	0.304	2.089	2.751	-0.374	0.456	0.318	0.000	0.544	-0.058
	40	0.476	0.309	0.278	0.151	0.567	-0.123	0.121	0.896	-0.090	0.613	0.036	0.577
	80	0.522	0.362	0.405	-0.021	0.484	0.413	0.358	0.036	0.596	0.124	0.600	-0.070
	160	0.341	0.060	0.540	0.244	0.313	0.268	-0.150	-0.361	0.433	0.147	0.401	0.387
	320	0.403	0.127	0.523	0.239	0.128	-0.052	0.448	0.018	0.338	-0.088	0.496	0.033
	640	0.818	0.574	0.850	0.716	0.289	0.141	0.402	-0.031	0.265	-0.103	0.433	0.050
	800	0.862	0.638	0.595	0.938	0.430	0.184	0.360	0.018	0.532	0.477	0.465	0.092
	1280	0.806	0.744	0.539	1.031	0.628	0.259	0.515	0.164	0.604	0.284	0.456	0.515
Bias Rate	20	2.031	-1.624	7.845	5.976	3.666	5.650	-2.913	2.607	0.662	-1.007	4.043	-1.982
	40	1.654	0.606	4.299	1.339	3.047	-3.025	0.011	3.191	-1.704	2.591	-0.733	2.449
	80	1.119	0.290	0.714	-1.110	2.489	2.771	7.313	-0.202	3.149	0.566	1.546	-1.888
	160	1.393	-0.552	1.465	-0.191	3.252	0.708	-2.784	-6.186	1.080	-0.419	4.937	2.906
	320	0.973	-0.234	1.312	-0.382	-0.195	-1.475	1.244	-0.885	1.185	-1.535	3.168	-0.761
	640	1.240	-0.141	0.345	0.109	2.029	-0.103	2.750	-1.224	1.945	-1.755	2.419	-0.656
	800	1.020	-0.143	0.746	0.175	1.701	-0.328	2.072	-0.888	2.382	0.276	2.540	-0.398
	1280	0.905	-0.125	0.980	0.147	0.862	-0.182	2.358	-0.299	2.358	-0.167	3.177	0.209
MAE	20	0.048	2.841	0.121	5.297	0.037	2.825	0.041	3.544	0.036	5.351	0.075	5.839
	40	0.036	1.439	0.064	3.233	0.056	5.277	0.060	2.169	0.063	2.469	0.044	2.540
	80	0.024	1.204	0.027	1.639	0.054	3.448	0.090	7.017	0.044	4.237	0.036	3.815
	160	0.030	1.471	0.023	1.045	0.055	2.138	0.056	4.350	0.038	1.395	0.057	2.637
	320	0.022	0.938	0.020	0.914	0.034	2.490	0.031	2.062	0.037	1.899	0.048	2.000
	640	0.016	0.497	0.008	0.575	0.040	1.874	0.051	1.906	0.045	2.065	0.043	1.637
	800	0.013	0.499	0.012	0.500	0.033	1.231	0.042	1.604	0.031	1.111	0.040	1.342
	1280	0.012	0.479	0.013	0.460	0.018	1.081	0.034	1.301	0.031	0.933	0.036	0.868
Number of Convegence in 100 Trials (Log-likelihood)	20	30 (-10.31)				4 (-7.49)							
	40	52 (-21.09)				9 (-15.75)							
	80	63 (-43.13)				10 (-17.30)							
	160	72 (-86.93)				26 (-33.59)							
	320	79 (-177.63)				49 (-73.159)							
	640	81 (-354.22)				66 (-147.09)							
	800	92 (-444.29)				66 (-181.79)							
	1280	91 (-712.28)				76 (-294.70)							

Exhibit 3 Results of Monte Carlo Simulation

(2) $\beta i=0.05,\ \Delta i=2.5\ (i=1\sim4,\ \tau i=\Delta i/\beta i=50)$

a) 2 attributes

Test Statistics	Number of Obs.	β2	Δ2	β1	Δ1
Average of Estimated Parameters	20	0.174	8.070	0.281	11.199
	40	0.121	4.656	0.179	6.813
	80	0.126	4.250	0.080	3.666
	160	0.072	3.153	0.062	2.977
	320	0.060	2.826	0.056	2.738
	640	0.056	2.685	0.052	2.591
	800	0.055	2.615	0.051	2.562
	1280	0.053	2.570	0.051	2.547
Standard Error	20	0.253	9.600	0.411	13.271
	40	0.194	7.483	0.313	8.928
	80	0.295	5.570	0.124	4.006
	160	0.074	2.133	0.028	1.070
	320	0.024	0.921	0.020	0.670
	640	0.014	0.418	0.012	0.383
	800	0.011	0.344	0.010	0.326
	1280	0.008	0.246	0.008	0.255
t-Value	20	0.688	0.841	0.682	0.844
	40	0.621	0.622	0.571	0.763
	80	0.428	0.763	0.647	0.915
	160	0.965	1.478	2.193	2.782
	320	2.479	3.069	2.826	4.090
	640	4.133	6.428	4.457	6.773
	800	4.798	7.607	5.128	7.847
	1280	6.598	10.460	6.416	10.000
Bias Rate	20	2.486	2.228	4.613	3.480
	40	1.413	0.862	2.573	1.725
	80	1.525	0.700	0.602	0.466
	160	0.437	0.261	0.237	0.191
	320	0.200	0.130	0.119	0.095
	640	0.128	0.074	0.039	0.037
	800	0.093	0.046	0.022	0.025
	1280	0.058	0.028	0.017	0.019
MAE	20	0.145	6.485	0.249	9.710
	40	0.086	3.587	0.143	4.870
	80	0.086	2.288	0.043	1.607
	160	0.031	1.079	0.021	0.824
	320	0.016	0.554	0.014	0.518
	640	0.011	0.337	0.009	0.309
	800	0.009	0.283	0.008	0.266
	1280	0.006	0.211	0.006	0.201

b) 4 attributes

Test Statistics	Number of Obs.	β4	Δ4	β3	Δ3	β2	Δ2	β1	Δ1
Average of Estimated Parameters	20	0.071	3.454	0.099	5.756	0.589	9.078	0.052	1.764
	40	0.098	6.052	0.082	5.779	0.057	6.138	0.071	6.135
	80	0.110	7.259	0.093	4.968	0.091	5.159	0.111	5.210
	160	0.166	6.402	0.069	2.974	0.033	-1.622	0.082	3.711
	320	0.065	3.538	0.095	3.546	0.085	3.248	0.099	3.817
	640	0.063	2.857	0.079	3.059	0.073	3.164	0.065	2.867
	800	0.060	2.824	0.058	2.646	0.069	2.888	0.066	2.872
	1280	0.059	2.880	0.059	2.688	0.063	2.852	0.062	2.727
Standard Error	20	0.062	2.686	0.159	5.355	0.847	6.994	0.079	3.247
	40	0.153	4.305	0.112	4.623	0.054	4.254	0.076	4.256
	80	0.167	9.162	0.211	6.554	0.131	2.672	0.259	12.185
	160	0.306	7.396	0.076	2.757	0.361	24.339	0.135	5.047
	320	0.058	3.319	0.081	2.863	0.090	2.805	0.125	3.943
	640	0.053	1.063	0.084	2.193	0.044	1.501	0.031	0.965
	800	0.029	0.767	0.026	0.796	0.029	1.143	0.064	1.343
	1280	0.027	0.705	0.033	0.855	0.025	0.617	0.023	0.605
t-Value	20	1.137	1.286	0.621	1.075	0.695	1.298	0.656	0.543
	40	0.640	1.406	0.730	1.250	1.067	1.443	0.933	1.442
	80	0.660	0.792	0.440	0.758	0.692	1.931	0.429	0.428
	160	0.542	0.866	0.909	1.079	0.092	-0.067	0.606	0.735
	320	1.126	1.066	1.166	1.238	0.936	1.158	0.792	0.968
	640	1.182	2.687	0.948	1.395	1.671	2.107	2.081	2.972
	800	2.089	3.684	2.275	3.323	2.378	2.528	1.036	2.139
	1280	2.174	4.084	1.783	3.145	2.550	4.620	2.721	4.505
Bias Rate	20	0.411	0.381	0.972	1.302	10.784	2.631	0.031	-0.294
	40	0.963	1.421	0.632	1.312	0.148	1.455	0.419	1.454
	80	1.205	1.904	0.856	0.987	0.820	1.064	1.221	1.084
	160	2.318	1.561	0.385	0.190	-0.338	-1.649	0.633	0.484
	320	0.296	0.415	0.896	0.418	0.692	0.299	0.980	0.527
	640	0.256	0.143	0.584	0.224	0.468	0.266	0.302	0.147
	800	0.209	0.130	0.164	0.058	0.379	0.155	0.322	0.149
	1280	0.180	0.152	0.181	0.075	0.256	0.141	0.234	0.091
MAE	20	0.050	2.424	0.094	4.035	0.559	6.578	0.058	2.576
	40	0.104	3.613	0.070	3.806	0.035	3.700	0.063	3.696
	80	0.105	4.801	0.103	4.799	0.078	2.798	0.130	7.073
	160	0.136	4.593	0.046	2.180	0.166	9.318	0.054	2.530
	320	0.028	1.480	0.050	1.694	0.046	1.298	0.062	1.902
	640	0.025	0.729	0.039	0.996	0.033	1.086	0.025	0.727
	800	0.022	0.635	0.018	0.557	0.025	0.833	0.027	0.745
	1280	0.017	0.543	0.019	0.487	0.020	0.556	0.017	0.466

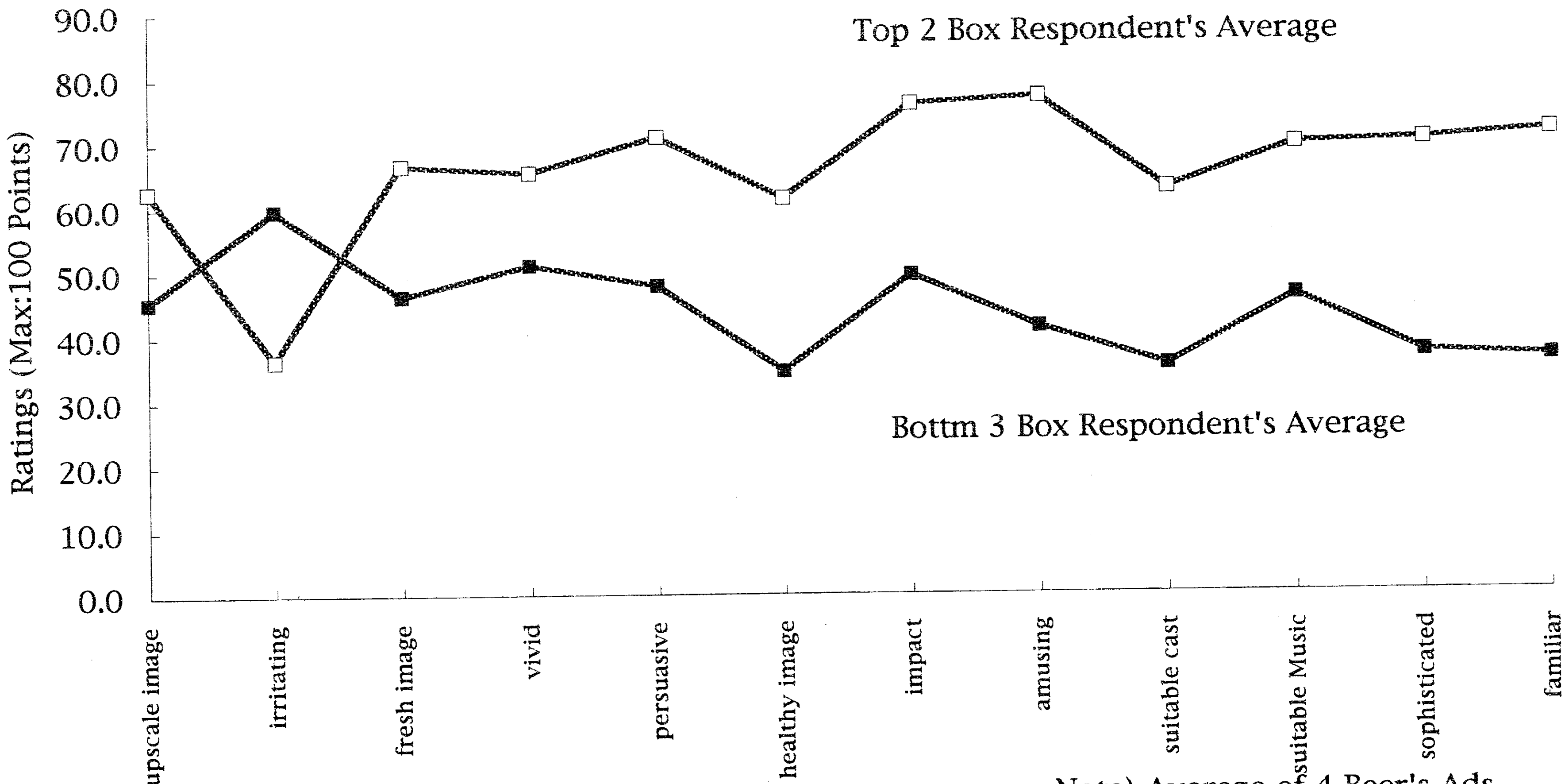
Test Statistics	Number of Obs.	a) 2 attributes	b) 4 attributes
Number of Convegence in 100 Trials (Log-likelihood)	20	36 (-7.65)	4 (-12.83)
	40	65 (-16.61)	15 (-27.79)
	80	82 (-32.83)	16 (-29.03)
	160	96 (-67.10)	14 (-30.46)
	320	100 (-136.65)	39 (-56.64)
	640	99 (-276.22)	63 (-115.28)
	800	100 (-347.38)	56 (-145.18)
	1280	100 (-557.92)	79 (-230.15)

Exhibit 3 Results of Monte Carlo Simulation

(3) $\beta i=0.1$, $\Delta i=5.0$ ($i=1\sim4$, $\tau i=\Delta i/\beta i=50$)

Test Statistics	Number of Obs.	a) 2 attributes $\beta 2$	a) $\Delta 2$	a) $\beta 1$	a) $\Delta 1$	b) 4 attributes $\beta 4$	b) $\Delta 4$	b) $\beta 3$	b) $\Delta 3$	b) $\beta 2$	b) $\Delta 2$	b) $\beta 1$	b) $\Delta 1$
Average of Estimated Parameters	20	0.203	8.346	0.300	13.506	0.124	6.884	-0.004	-0.495	0.083	7.876	0.495	9.709
	40	0.218	10.346	0.246	11.542	0.100	6.225	0.090	7.160	0.000	-1.733	0.062	5.737
	80	0.146	7.138	0.156	7.605	0.084	6.619	0.122	9.067	0.230	12.121	0.121	6.815
	160	0.129	6.265	0.123	6.108	0.038	1.899	0.105	7.516	0.034	-0.851	0.064	5.345
	320	0.113	5.597	0.110	5.502	0.145	6.818	0.140	6.522	0.129	5.216	0.111	4.124
	640	0.107	5.292	0.104	5.211	0.115	5.485	0.130	6.271	0.140	6.583	0.106	5.299
	800	0.105	5.205	0.102	5.100	0.113	5.583	0.127	6.100	0.120	5.724	0.114	5.492
	1280	0.104	5.136	0.100	5.028	0.109	5.287	0.100	4.981	0.101	4.920	0.106	5.112
Standard Error	20	0.362	23.178	0.267	10.640	0.124	4.811	0.292	19.513	0.140	5.762	0.997	9.382
	40	0.173	8.476	0.283	11.702	0.141	4.214	0.083	4.884	0.305	25.951	0.058	4.638
	80	0.110	4.985	0.138	6.253	0.079	7.130	0.109	6.092	0.265	8.340	0.094	5.392
	160	0.071	3.165	0.074	3.002	0.068	7.559	0.061	2.496	0.126	9.507	0.036	4.280
	320	0.031	1.326	0.029	1.298	0.092	3.987	0.100	3.629	0.100	6.999	0.122	10.022
	640	0.018	0.779	0.015	0.721	0.058	2.310	0.072	3.104	0.065	2.642	0.046	1.986
	800	0.013	0.623	0.013	0.619	0.038	1.795	0.077	3.160	0.054	2.141	0.050	1.763
	1280	0.010	0.475	0.009	0.407	0.075	2.855	0.036	1.823	0.033	1.559	0.053	2.099
t-Value	20	0.562	0.360	1.123	1.269	0.996	1.431	-0.014	-0.025	0.594	1.367	0.496	1.035
	40	1.258	1.221	0.869	0.986	0.712	1.477	1.077	1.466	0.000	-0.067	1.068	1.237
	80	1.335	1.432	1.130	1.216	1.069	0.928	1.123	1.488	0.869	1.453	1.276	1.264
	160	1.799	1.979	1.671	2.035	0.563	0.251	1.708	3.011	0.274	-0.090	1.771	1.249
	320	3.652	4.222	3.821	4.238	1.580	1.710	1.400	1.797	1.283	0.745	0.907	0.411
	640	6.069	6.793	6.923	7.224	1.985	2.374	1.803	2.020	2.161	2.492	2.314	2.669
	800	7.789	8.354	7.966	8.244	2.949	3.110	1.642	1.930	2.233	2.674	2.310	3.116
	1280	10.034	10.823	11.435	12.345	1.457	1.852	2.736	2.733	3.048	3.156	2.017	2.436
Bias Rate	20	1.033	0.669	1.997	1.701	0.240	0.377	-1.040	-1.099	-0.167	0.575	3.945	0.942
	40	1.180	1.069	1.458	1.308	0.003	0.245	-0.102	0.432	-0.999	-1.347	-0.379	0.147
	80	0.462	0.428	0.555	0.521	-0.156	0.324	0.225	0.813	1.298	1.424	0.205	0.363
	160	0.285	0.253	0.234	0.222	-0.619	-0.620	0.048	0.503	-0.655	-1.170	-0.358	0.069
	320	0.134	0.119	0.100	0.100	0.446	0.364	0.402	0.304	0.287	0.043	0.105	-0.175
	640	0.067	0.058	0.039	0.042	0.154	0.097	0.298	0.254	0.403	0.317	0.057	0.060
	800	0.048	0.041	0.018	0.020	0.135	0.117	0.270	0.220	0.203	0.145	0.145	0.098
	1280	0.036	0.027	0.003	0.006	0.086	0.057	-0.003	-0.004	0.010	-0.016	0.063	0.022
MAE	20	0.253	14.086	0.214	9.334	0.085	3.021	0.170	9.471	0.105	4.013	0.456	5.846
	40	0.133	6.193	0.165	7.314	0.092	2.517	0.062	3.443	0.142	10.475	0.059	3.005
	80	0.063	2.957	0.074	3.473	0.062	4.398	0.073	4.282	0.167	7.121	0.067	3.754
	160	0.040	1.850	0.038	1.735	0.062	3.872	0.052	2.516	0.096	6.623	0.045	3.125
	320	0.023	0.987	0.023	1.000	0.067	2.658	0.065	2.567	0.063	3.487	0.062	3.611
	640	0.014	0.606	0.012	0.577	0.035	1.445	0.045	1.907	0.050	1.988	0.030	1.125
	800	0.011	0.500	0.010	0.477	0.031	1.402	0.038	1.590	0.034	1.340	0.032	1.215
	1280	0.008	0.378	0.007	0.333	0.035	1.407	0.028	1.276	0.024	1.062	0.032	1.328
Number of Convegence in 100 Trials (Log-likelihood)	20	33 (-5.44)				7 (-28.15)							
	40	61 (-9.31)				10 (-19.13)							
	80	81 (-20.94)				13 (-34.39)							
	160	96 (-43.22)				5 (-39.23)							
	320	99 (-88.359)				30 (-43.197)							
	640	100 (-179.18)				38 (-78.679)							
	800	100 (-226.22)				51 (-100.06)							
	1280	100 (-364.61)				53 (-168.71)							

Exhibit 4 Average Ratings of Top 2 Box and Bottom 3 Box Respondents



Note) Averege of 4 Beer's Ads.

Exhibit 5 Result of Estimation (4 Beer Ad's)

	4 Variables			3 Variables			2 Variables		
	β	Δ	$T=\Delta/\beta^*$	β	Δ	$T=\Delta/\beta^*$	β	Δ	$T=\Delta/\beta^*$
familiar	0.089	5.365	60.35	0.088	5.190	58.84	0.090	5.339	59.65
	(0.035)	(1.903)		(0.032)	(1.849)		(0.042)	(2.081)	
amusing	0.088	5.712	65.28	0.089	5.925	66.50	0.086	5.972	69.76
	(0.030)	(1.957)		(0.027)	(1.966)		(0.022)	(1.718)	
suitable casting	-0.128	-24.032	187.31	0.055	1.512	27.34			
	(7.552)	(744.978)		(0.033)	(2.571)				
irritating	-0.105	-8.194	78.41						
	(0.072)	(5.385)							
Log-Likelihood	-52.58			-54.16			-54.95		

Standard error in parenthesis.

* Threshold T is given by Δ/β.

Exhibit 6 Estimated Thresholds of 2 Attributes

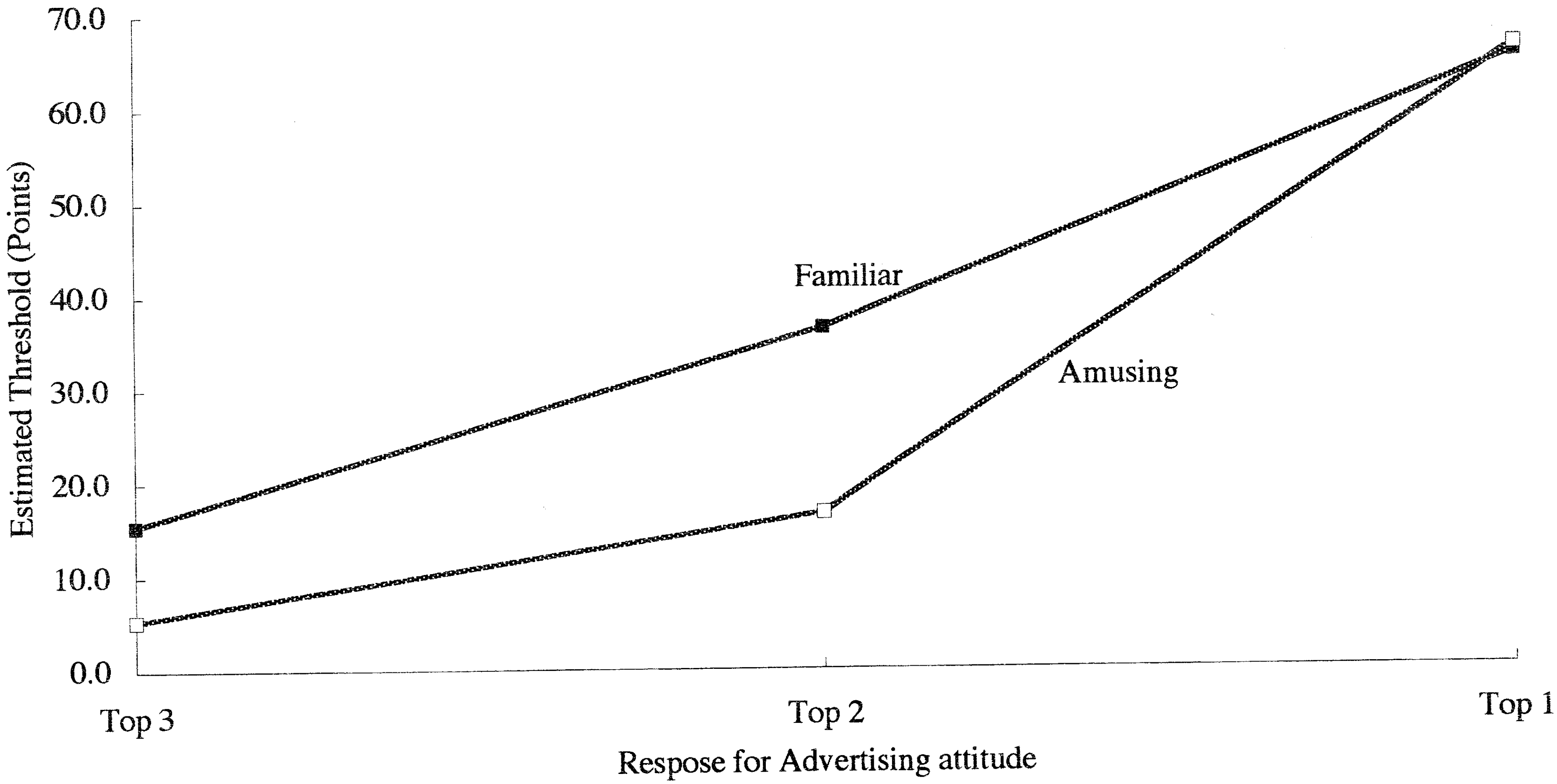


Exhibit 7-1 Probability Response Curve for Attribute "Familiar"

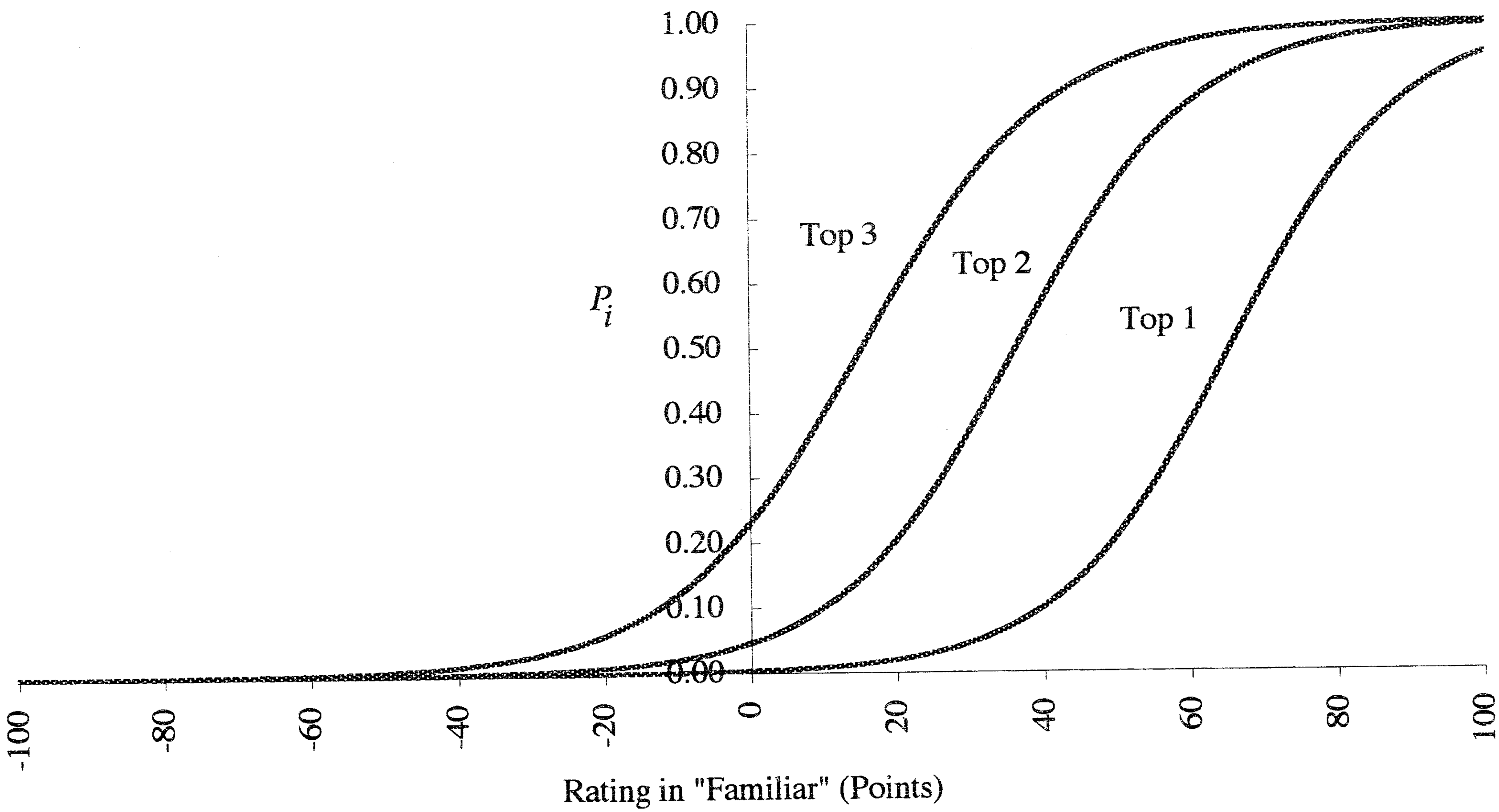


Exhibit 7-2 Probability Response Curve for Attribute "Amusing"

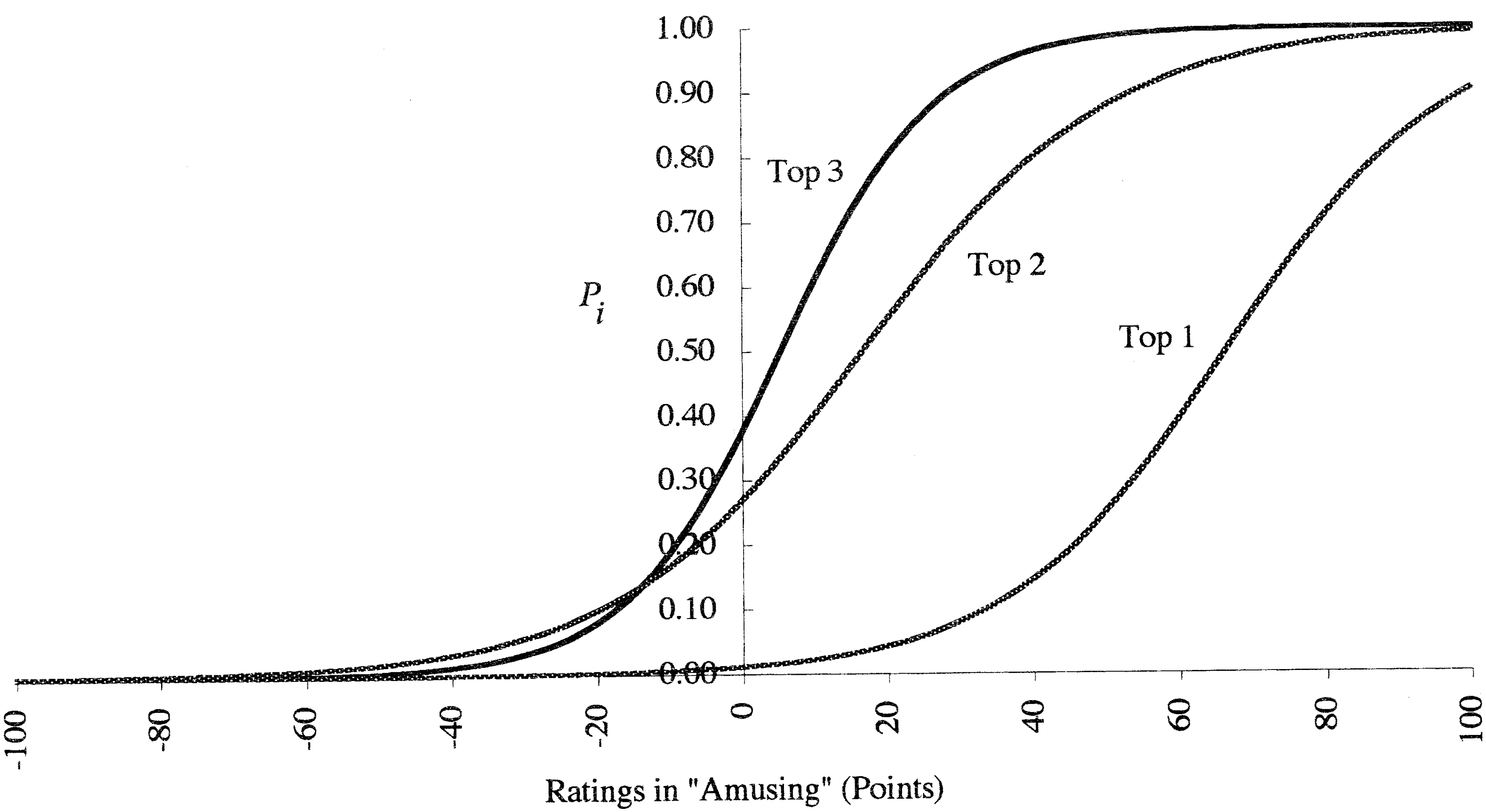


Exhibit 8 Results of Holdout Sample Tests
Holdout : "Ichiban Shibori's Ad"

Variables used*	Observed** %	Binomial Logit		Non-Compensatory Logit	
		LL	Predicted %	LL	Predicted %
x_2, x_8, x_9, x_{12}	45.7	-47.52	62.0	-52.58	58.9
x_8, x_9, x_{12}	44.7	-51.43	60.8	-54.16	58.2
x_8, x_{12}	44.7	-54.05	59.6	-54.95	57.6

*: x_2: irritating, x_8: amusing, x_9: suitable casting, x_{12}: familiar

**:Percentage of respondents who rated top 1 box(I like this ad very much). Observed percentage is slightly different due to missing values.