A MULTI-BRAND MULTI-SEGMENT MODEL OF CONSUMER BEHAVIOR: ON ESTIMATING THE PARAMETERS

John C. Bieda, Procter & Gamble

Introduction

The purpose of this paper is to report on the findings from an empirical study of the Minimum Chi Square estimation procedure when this estimation procedure is used for estimating the parameters of a heterogeneous linear learning model in the context of consumer purchasing behavior.

In 1965, Massy [1] proposed that the Minimum Chi Square estimation procedure be used to estimate the parameters of the homogeneous two operator linear learning model. Massy was able to show that the expected proportion of families that would purchase a given sequence of brands is a function of the parameters of the two operator model and the raw moments of the initial distribution of the probabilities of purchasing the two brands.

In addition, Massy suggested using data from continuous purchase diary panels to obtain the observed proportion of families that purchased a particular sequence of brands. These are the two pieces of data that are necessary for the Minimum Chi Square estimation procedure as can be seen by formula

$$X^2 = M \sum_{i=1}^{N} \frac{(P_i - \pi_i\,(\delta))^2}{\pi_i\,(\delta)} .$$

where M is the sample size for the panel; N is the length of the purchase string; P_i is the observed proportion of families that have purchased string i; π_i is the expected proportion of families that will purchase string i; and δ is the vector of parameters which included the parameters of the linear learning model and the parameters of the initial distribution of purchasing the brands.

Massy used Crámer's proof, that when the above expression is at a minimum the estimated parameters are asymptotically equivalent to maximum likelihood estimates, to justify using this procedure. Massy also pointed out that when the above expression is at the minimum the resulting Chi Square value can be used to test the fit of the model against the null hypothesis that the model does fit the observed data.

In a forthcoming doctoral dissertation, Bieda has expanded the two operator linear learning model to allow for heterogeneity among the population with respect to the model parameters. In a manner similar to that used by Massy, he has been able to show that the expected proportion of families that will purchase a particular sequence of brands is a function of the raw moments of the initial distribution of purchasing the brands, the weights associated with segments of the population that are assumed to be homogeneous with respect to a particular two operator linear learning model and the parameters of the learning models associated with each segment. Thus, Bieda was able to adopt Massy's estimation procedure.

With this background in mind, we now shall take a closer look at the estimation procedure. To minimize the above expression, we would ordinarily take the partial derivatives with respect to each of the parameters, set them equal to zero and solve the resulting set of equations. However, for both the homogeneous and heterogeneous models, the partical derivatives are highly non-linear and very complicated. Thus, there is a need to resort to some type of non-linear programming algorithm to solve for the minimum.

The algorithm that has been used for the homogeneous two operator in several studies that have been reported in the literature is Pattern Search. Time does not permit a discussion of this non-linear programming procedure; instead, I will simply refer you to an excellent discussion of it in Wilde's book, Optimum Seeking Methods. [2]

The fact that we have to use a non-linear programming procedure to minimize the Chi Square expression is the basis for the present study. It would seem that unless one can arrive at the true minimum - 1) one has no idea of what statistical properties the estimates have and 2) using the Chi Square value obtained to test the fit of the model would be incorrect since Cramer's proof requires that the Chi Square expression be at the minimum for the observed set of data.

In this study our primary research question then deals with the ability of the non-linear programming, Pattern Search, to arrive at the true minimum.

The Experiment

The design of the experiment for examining this question was factorial in nature. Four formulations of the heterogeneous two operator learning model were examined. These were 1) a Two Beta Equal Lambda Model, 2) a One Beta Equal Lambda Model, 3) a Two Beta, Unequal Lambda Model and 4) a One Beta Unequal Lambda Model.

Here the unequal and equal lambdas refer to whether or not the slopes of the operators are unequal or equal and the one and two betas refer to whether the initial distribution of probabilities is specified as a single beta distribution or a sum of weighted beta distributions.

The procedure used was as follows: 1) For a given formulation a set of parameters was arbitrarily chosen. 2) These parameters were used to obtain a set of exact proportions for purchase strings of length five using the model. 3) The exact proportions were submitted to the non-linear programming algorithm as if they were observed proportions from a consumer panel with another set of parameters used as starting values. 4) The non-linear programming algorithm was allowed to run until there was for all practical purposes no further improvement possible.

For each formulation of the model a set of parameters was chosen for a 1, 2 and 3 segment market. Thus, in all, twelve sets of exact proportions were generated.

Before going into the results, a word needs to be said about the starting values. Pattern Search, like many non-linear algorithms, must be given an initial vector of feasible starting parameters. We used the following procedure to select the starting values. Regardless of how many segments were actually used to generate the exact proportions, the exact proportions were first submitted to a one segment version of the model. The ending parameters were then duplicated, the segment weight cut in half and these values served as the starting values for a two segment version of the same model. If the exact data was generated with a three segment model the ending two segment parameters served as the starting values for the three segment version. Here, however, we took the larger of the two segments duplicated the parameters and split the larger segment weight in half.

In several cases, an alternate procedure was also employed. Here we simply duplicated our initial starting values for the one segment version for say exact data generated for a two segment market, split the segment weight in half and used these values as starting values. We will have more to say about the starting values in the discussion section.

Results

In Table 1, we have presented the Chi Square values that were obtained using the stepwise starting value procedure and those obtained for the selected cases where the second set of starting values were used.

For the equal lambda model the observed Chi Square values were generally quite low when the stepwise starting values were used.

For the two beta formulation the one and two segments observed χ^2's were $.55 \times 10^{-5}$ and $.17 \times 10^{-4}$ respectively. For the one beta formulation the one and three segment cases they were $.49 \times 10^{-7}$ and $.26 \times 10^{-5}$ respectively. The two exceptions to the low observed Chi Square value for the equal lambda formulation were for the two beta 3 segment and one beta 2 segment cases with the values of $.30 \times 10^{-1}$ and $.11 \times 10^{-3}$. However, with the alternative set of starting values both of the observed values dropped to levels comparable with the others $.87 \times 10^{-4}$ for the two beta 3 segment cases and $.29 \times 10^{-4}$ for the one beta 2 segment formulation.

With the unequal lambda model the results are somewhat different. Here we see that, for both the two beta and the one beta formulations, as the number of segments increases, the observed Chi Square steadily decreases. Although not shown in the table using alternative starting values did not improve any of the values.

In Tables 2 through 4, we have presented the estimated parameters for the twelve sets of exact data. Under each of the estimated parameters, in parentheses, is the exact parameter being estimated.

Looking first at Table 2. Here it can be seen that for the Two Beta Equal Lambda Model the estimated parameters are very close to the exact parameters in almost every case. The exception involves the α_{11} and λ for the three segment market. In addition, the beta parameters, while estimating the mean of the initial distribution quite closely, are not estimating the second or higher raw moments very well.

Turning now to Table 3. The estimates for the two brand one beta equal lambda formulation are almost precisely the exact parameters for the one and two segment markets. The beta parameters for the one segment market are also very close, however, for the two segment market they are underestimated. For the three segment market the model parameters are generally close with the exception of the segment weights. The beta parameters, however, are badly overestimated.

In Table 4, it can be seen here that as more segments are added the estimated parameters are closer to the exact parameters. In the three segment case the beta parameters are very close to those used to generate the data, the model parameters in almost every case are also quite close to the exact parameters.

In Table 5, we have a virtual repeat of Table 4. As more segments are added the estimated parameters are closer to the exact parameters. The exception here is in the estimates of the beta parameter. The estimates for the two segment market are closer to the exact beta parameters than the estimates for the three segment market.

Discussion

It will be noted that in these experiments we did not introduce any error into the exact proportion generated from the known set of parameters. The reason for this is quite simple. We are trying to determine if for a given set of observed proportions we can move to the global minimum using Pattern Search and the model that generated the proportions. If we can, then we know from Crámer's proof that the estimated parameters are equivalent to the maximum likelihood estimates.

The results are somewhat mixed in the equal lambda case, the stepwise starting procedure generally provided very low observed values. However, using this procedure a local minima was definitely encountered in the two beta 3 segment case.

For the unequal lambda case a local minima was encountered in both the one and two segment cases. Even when an alternative set of starting values was used other local minima were encountered. Only in the case of the three segment market were the results reasonably close.

In general, the estimates of the beta parameters, for both the equal and unequal formulations, were quite poor.

It should be noted that in no case did the Chi Square value ever reach zero. For these particular experiments, it should have been possible to reach zero since the exact proportions were carried to sixteen decimal places. One explanation besides local minima for why smaller values were not observed lies in round off error in the calculations; another is that the step size in the search routine was not allowed to decrease to a small enough value to obtain the desired accuracy.

Conclusion

Generally speaking, we were able to get quite close to the global minima. However, in several cases the observed Chi Square values were sufficiently far enough away from the known true minima, 0.0, to cast some doubt on the ability of the search routine to reach the global minima in all instances. Particularly disturbing was the fact that when an alternative set of parameters was used for the unequal lambda formation, no improvement was found in the observed Chi Square values.

The general conclusion seems to be that a number of starting values should be used even though one is not guaranteed of reaching the global minima using Pattern Search for estimating the parameters of a multi-brand multi-segment linear learning model.

REFERENCES

(1) Massy, William F., "Estimation of Parameters for the Linear Learning Model," Working Paper #78, 1965.

(2) Wilde, D. J., Optimum Seeking Methods, Prentice Hall, Inc. 1964.

TABLE 1

Observed Chi Square Value for Stepwise and Selected Second Set Starting Values

Date Generated With		Observed X^2 Values Stepwise Starting Values Used	X^2 Values, Second Set of Starting Values Used
2 Beta, Equal	1 Segment	$.005528 \times 10^{-3}$	
	2 Segment	$.017580 \times 10^{-3}$	
	3 Segment	30.035×10^{-3}	$.08711 \times 10^{-3}$
1 Beta, Equal	1 Segment	$.000049 \times 10^{-3}$	
	2 Segment	$.1076 \times 10^{-3}$	$.02871 \times 10^{-3}$
	3 Segment	$.00261 \times 10^{-3}$	
2 Beta Unequal	1 Segment	4.397×10^{-3}	
	2 Segment	$.2880 \times 10^{-3}$	
	3 Segment	$.0159 \times 10^{-3}$	
1 Beta Unequal	1 Segment	4.225×10^{-3}	
	2 Segment	$.5668 \times 10^{-3}$	
	3 Segment	$.004082 \times 10^{-3}$	

TABLE 2

Estimated and Exact Parameters Two Brand Two Beta Equal Lambda Formulation

Exact Parameters Being Estimated in Parentheses

	One Segment Market	Two Segment Market		Three Segment Market		
	Seg. 1	Seg. 1	Seg. 2	Seg. 1	Seg. 2	Seg. 3
α_{11}	0.3145 (0.3145)	0.364 (0.3145)	0.363 (.382)	0.30 (0.382)	0.21 (0.3145)	0.51 (0.4468)
λ	0.644 (0.644)	0.621 (0.644)	0.621 (0.618)	0.69 (0.618)	0.79 (0.644)	0.47 (0.54)
α_{12}	0.004 (0.004)	0.19 (0.004)	0.18 (0.029)	0.0007 (0.029)	0.04 (0.035)	0.03 (0.0282)
Segment Weight		0.41 (0.40)	0.59 (0.60)	0.31 (0.30)	0.12 (0.20)	0.57 (0.50)
1st Parameter 1st Beta Distribution	4.18 (22.50)	11.87 (22.50)		2.36 (22.50)		
2nd Parameter 1st Beta Distribution	1.26 (6.04)	2.16 (6.04)		2.40 (6.04)		
Weight 1st Beta Distribution	0.09 (0.10)	0.07 (0.10)		0.27 (0.10)		
1st Parameter 2nd Beta Distribution	3.97 (6.04)	4.99 (6.04)		7.80 (6.04)		
2nd Parameter 2nd Beta Distribution	14.16 (22.50)	17.13 (22.50)		34.64 (22.50)		

TABLE 3

Estimated and Exact Parameters For the
Two Brand One Beta Equal Lambda Formulation

Exact Parameters Being Estimated
In Parentheses

	One Segment Market	Two Segment Market		Three Segment Market		
	Seg. 1	Seg. 1	Seg. 2	Seg. 1	Seg. 2	Seg. 3
α_{11}	0.262 (0.262)	0.24 (0.262)	0.40 (0.41)	0.264 (0.262)	0.38 (0.3135)	0.40 (0.41)
λ	0.683 (0.683)	0.70 (0.683)	0.60 (0.59)	0.622 (0.683)	0.60 (0.557)	0.59 (0.59)
α_{12}	0.0089 (0.0089)	0.01 (0.0089)	0.03 (0.025)	0.0098 (0.0089)	0.024 (0.0165)	0.024 (0.025)
Segment Weight		0.38 (0.40)	0.64 (0.60)	0.44 (0.30)	0.48 (0.30)	0.08 (0.40)
1st Beta Parameter	0.90 (0.899)	4.18 (5.82)		8.42 (13.62)		
2nd Beta Parameter	8.55 (8.50)	8.18 (12.00)		38.50 (62.37)		

TABLE 4

Estimated and Exact Parameters for the Two Brand
Two Beta Unequal Lambda Formulation

Exact Parameters Being Estimated
in Parentheses

	One Segment Market	Two Segment Market		Three Segment Market		
	Seg. 1	Seg. 1	Seg. 2	Seg. 1	Seg. 2	Seg. 3
α_{11}	0.22 (.32)	0.19 (.10)	0.18 (.32)	0.20 (.25)	0.28 (.32)	0.27 (.10)
λ_1	0.78 (.65)	0.79 (.85)	0.79 (.65)	0.61 (.50)	0.70 (.65)	0.70 (.85)
α_{12}	0.03 (.03)	0.05 (.05)	0.04 (.03)	0.02 (.10)	0.03 (.03)	0.03 (.05)
λ_2	0.86 (.80)	0.72 (.70)	0.71 (.80)	0.60 (.60)	0.79 (.80)	0.79 (.70)
Segment Weight		0.53 (.60)	0.47 (.40)	0.50 (.40)	0.30 (.40)	0.20 (.20)
1st Parameter 1st Beta Distribution	6.63 (10.50)	28.03 (10.50)		6.58 (10.50)		
2nd Parameter 1st Beta Distribution	4.40 (4.55)	12.89 (4.55)		3.43 (4.55)		
Weight 1st Beta Distribution	0.31 (.20)	0.21 (.20)		0.22 (.20)		
1st Parameter 2nd Beta Distribution	4.03 (7.35)	11.36 (7.35)		7.40 (7.35)		
2nd Parameter 2nd Beta Distribution	57.15 (59.20)	95.14 (59.20)		62.76 (59.20)		

TABLE 5

Estimated and Exact Parameters Two Brand
One Beta Unequal Lambda Formulation

Exact Parameters Being Estimated
in Parentheses

	One Segment Market	Two Segment Market		Three Segment Market		
	Seg. 1	Seg. 1	Seg. 2	Seg. 1	Seg. 2	Seg. 3
α_{11}	0.24 (0.32)	0.19 (0.10)	0.19 (0.32)	0.14 (0.10)	0.25 (0.25)	0.31 (0.32)
λ_1	0.76 (0.65)	0.81 (0.85)	0.81 (0.65)	0.70 (0.85)	0.57 (0.50)	0.68 (0.65)
α_{12}	0.03 (0.03)	0.05 (0.05)	0.05 (0.03)	0.03 (0.05)	0.02 (0.01)	0.03 (0.03)
λ_2	0.87 (0.80)	0.69 (0.70)	0.69 (0.80)	0.79 (0.70)	0.59 (0.60)	0.82 (0.80)
Segment Weight		0.56 (0.60)	0.56 (0.40)	0.18 (0.20)	0.51 (0.40)	0.31 (0.40)
1st Beta Parameter	1.41 (6.04)	5.70 (6.04)		10.27 (6.04)		
2nd Beta Parameter	5.09 (22.05)	21.19 (22.05)		38.28 (22.05)		