A CROSS VALIDATION STUDY COMPARING TWO APPROACHES TO FITTING REGRESSION MODELS IN A MIXTURE OF TWO POPULATIONS: CLASSICAL METHODS VS. THE MIXTURE MODEL

Steven M. LaLonde, Rochester Institute of Technology
98 Lomb Memorial Drive, Rochester, NY 14623-5604

KEY WORDS: mixture regression, cluster analysis, customer survey, conjoint analysis.

1. INTRODUCTION

Designed experiments in product features are typically applied to new product design problems to determine which features are important to product attractiveness. This is often referred to as a "conjoint analysis" in the marketing research literature. Each respondent is exposed to several product profiles, combinations of product features, and asked to rank, or rate the attractiveness of the profiles. Linear models are often fit to the aggregate data in order to derive the 'importance' of individual product features. Often times individual models for respondents are fit, and the resulting regression coefficients are grouped using a post-hoc clustering algorithm in order to explore segmentation.

Latent class conjoint analysis, based on a mixture model, combines the two analysis objectives, conjoint analysis and segmentation, into one step. Latent class conjoint analysis fits regression equations to classes of respondents exhibiting similar response patterns. The result is a number of customer segments, each with its own product feature preferences.

This paper compares the results from the usual conjoint analysis with a latent class conjoint analysis of customer ratings of product profiles of banking services. In addition, the stability of the two approaches is investigated.

2. THE PROBLEM DESCRIPTION

This paper will focus on the analysis of data based on a survey of five hundred forty two commercial banking customers. The data has been altered from the original study and may not represent the total population of customers. The data was collected in a two-stage telephone interview with individual customers. During the first phone contact with customers participation was elicited, and a few preliminary questions were asked. Next, a package of product profiles were distributed by mail to the customers with directions for rating the product profiles enclosed. In a second phone call the ratings of the product profiles were collected, along with some additional survey information. The customers were paid an incentive upon completion of all requirements for the study.

Customers were asked to rate their likelihood of purchase on sixteen product profiles on a five-point scale where 1 represented 'Not at all likely to purchase' and 5 represented 'Extremely likely to purchase.' The sixteen product profiles were formed using an orthogonal design of five factors, each with two levels, of the following product features:

1. Number of Transactions (100 per month versus unlimited)
2. Overdraft Protection (none versus $5,000)
3. Financial Advisor (from a pool versus a personally assigned advisor)
4. Personal Banking Package (not included versus included with the Commercial Banking Package)
5. Price ($40 per month versus $80 per month)

The objective of the analysis is to determine which of these features are most important in determining the overall package attractiveness.

In addition, it is believed that some commercial customers are more price-sensitive than others. Understanding the needs of heterogeneous customer segments would allow a more tailored marketing program (product/ pricing/ channel/ communications). Estimating the size of the less price-sensitive segment of commercial customers is particularly of interest.

3. ISSUES TO BE CONSIDERED

For the purpose of this investigation we will assume that the sample we are working with has no obvious a priori segmentation. In the more general situation, there may be some a priori segmentation; based on the size of the account, or the type of business, or some other demographic variable.

In addition, this analysis is based on ratings of the product profiles, whereas the data collected is often times rankings of those same profiles. Rankings force the customer to differentiate between each and every product profile, whereas the same rating can be assigned to multiple product profiles. This paper does not address the issue of rankings versus ratings.

This type of problem is also approached using a discrete choice exercise, where multiple product profiles are presented to the customer in sets, and the customer chooses one among the group of profiles in each set. Latent class (mixture) analysis can also be applied to data collected from a discrete choice exercise as well. This paper does not address the issue of conjoint ratings versus the discrete choice approach.

3. AGGREGATE ANALYSIS

Simply submitting all 542 respondents to a multiple regression procedure forms the aggregate analysis. The resulting coefficients for each of the variables are show in Table 1. Price, with a negative coefficient (-1.35) is the most important variable, followed by Number of Transactions (.34), Overdraft Protection (.33), and Personal Banking Package (.26). The Financial Advisor (.05) appears not to be important.

5. TRADITIONAL SEGMENTATION

In the traditional segmentation approach, individual models were fit to each of the 542 respondents, with each model based the ratings of the 16 product profiles. The coefficients where then clustered into two mutually exclusive groups using Ward's algorithm. The average coefficients in each group were then calculated and are shown in Table 1. This analysis reveals a more price sensitive group one (-1.91), and less price sensitive group two (-.36). Furthermore, group one is less interested, relative to group two, in the three products identified in the aggregate analysis as important. Thirty-six percent (36%) of the customers fall into the less price sensitive group.

6. LATENT CLASS SEGMENTATION

In the latent class segmentation approach, the modeling and segmentation occurs simultaneously (DeSarbo et. al. 1992). The algorithm used here was written in SAS, primarily in PROC IML. The approach assigns respondents to the two segments in a way that maximizes the joint likelihood of the mixture distribution. The final model coefficients, and proportion of respondents in each group, can be calculated based on the resulting posterior probabilities, or on disjoint classifications based on the posterior probabilities. Table 1 shows the results from the latent class segmentation approach. The models resulting from the fuzzy posterior probabilities are identified in columns labeled as Latent Class 1, whereas the models from the disjoint classifications are in the columns labeled Latent Class 2. These two methods yield proportions for the less price sensitive segment of 38% and 37%, respectively.

Table 1: Preliminary Results

	Aggregate	Ward's		Latent Class 1		Latent Class 2	
Variable	Analysis	Cluster 1	Cluster 2	Cluster 1	Cluster 2	Cluster 1	Cluster 2
Number of Transactions	0.34	0.23	0.54	0.26	0.47	0.25	0.48
Overdraft Protection	0.33	0.27	0.43	0.34	0.31	0.34	0.32
Financial Advisor	0.05	0.06	0.03	0.09	-0.02	0.09	-0.01
Personal Banking Package	0.26	0.29	0.21	0.30	0.20	0.30	0.20
Price	-1.35	-1.91	-0.36	-1.91	-0.40	-1.92	-0.44
N	542	347	195	334	208	339	203
Proportion	1.00	.640	.360	.616	.384	.625	.375

7. COMPARISON OF THE TWO METHODS

On the surface it appears that the two methods produce very similar results. The coefficients for the product features for the two segments are very similar, as is the proportion of customers that are in the less price sensitive segment. Indeed, when classifications of individual customers are compared between the two methods, 91% of the customers fall into the same segment classifications. The classification is summarized in Table 2 and Table 3.

Table 2: Counts

		Latent Class 1	2	All
Traditional	1	320	27	348
(Ward's)	2	19	176	197
	All	340	205	542

Table 3: Proportion s

		Latent Class 1	2	All
Traditional	1	0.59	0.05	0.64
(Ward's)	2	0.04	0.32	0.36
	All	0.63	0.38	1.00

This measure of agreement is considerable higher than that found by Desarbo et. al. (1992). In that paper the authors found agreement between the latent class approach and a traditional approach based on a Ward's classification to be only 60%. It should be noted that in Desarbo's application, there were four segments to be classified, and hence much more opportunity for misclassification.

8. INVESTIGATION OF THE ROBUSTNESS OF THE EACH APPROACH

The robustness of the traditional approach will be investigated in two ways. First, other methods for clustering the coefficients will be computed, and their agreement with the latent class approach quantified. Second, resampling will be used to investigate the distributions of mean segment coefficients, as well as the distributions of the segment proportions.

Robustness of the latent class conjoint approach will also be investigated by investigating the distributions of mean segment coefficients, as well as the distributions of the segment proportions from resampling.

9. RESULTS FROM COMPARISON OF VARIOUS TRADIATIONAL CLUSTERING METHODS

Table 4 displays the results from the classification of the individual regression coefficients using four different clustering methods. The first method, using Ward's linkage, is the same as was shown in Table 1. The other three methods, single linkage, complete linkage, and average linkage, are shown for comparison. It appears upon examination of Table 2 that the results from the four methods have little in common. They have very different average coefficients for the variables, and very different cluster sizes.

The single linkage method is known to suffer from chaining, which often results in the anomaly seen here where the majority of the data is separated from small clusters of outliers. The results from the complete and average linkage methods yield more reasonable proportions, although both solutions result in large negative

Table 4: Results from Various Traditional Clustering Methods

	Ward's Linkage		Single Linkage		Complete Linkage		Average Linkage	
Variable	Cluster 1	Cluster 1	Cluster 1	Cluster 2	Cluster 1	Cluster 2	Cluster 1	Cluster 2
Number of Transactions	.23	.54	.34	.00	.44	-.36	.36	.10
Overdraft Protection	.27	.43	.33	-.50	.36	.10	.35	.09
Financial Advisor	.06	.03	.05	1.5	.11	-.36	.08	-.28
Personal Banking Package	.29	.21	.26	-.50	.33	-.22	.31	-.39
Price	-1.91	-0.36	-1.35	-1.00	-1.49	-.27	-1.49	.40
N	347	195	541	1	478	64	502	40
Proportion	.640	.360	.998	.002	.882	.118	.926	.074

coefficients for variables that would normally be expected to have positive weights.

It becomes evident here that I chose Ward's method for Table 1 in part because its results were closest to those from the latent class methodology. The coefficients from Ward's method are also the most interpretable, that is, variables that should have positive weights do, and visa versa.

10. RESAMPLING METHODOLOGY

The objective of the resampling is to compare the ability to reproduce the two segments statistics from samples from the 542 respondents in the total population. Five hundred forty two (542) samples of size one, with replacement, are drawn from the population at each replication. This is done two hundred twenty five (225) times for each statistical technique to be applied to the data.

Using the binomial distribution, and assuming that the split of the two segments was 38% and 62%, the resulting distribution would theoretically have a mean of 38% (for the less price sensitive segment), and a standard deviation of 2.09%, if the respondents could be classified without any error. Additional variability, or any offset, resulting from the classification methodologies can therefore be attributed to the methodologies and not the resampling.

11. RESULTS FROM RESAMPLING

Table 5 shows the results from two hundred twenty five (225) replications of resampling for each of the methodologies applied. From a practical perspective, methods that have minimum variance, and smaller ranges, would be preferred. In addition, methods that have less of an offset from the population would also be preferred.

The Fuzzy Latent Class method had a population proportion of .620 for Cluster 1, and a resampling average of .628, a relatively small offset of .008. In addition, the standard deviation of .0259 is very close to the theoretical sampling standard deviation of .0209. The two hundred twenty five replications varied from .555 to .692, a range of .137.

Table 5: Results from Resampling

	Cluster	N	Mean	Std Dev	Minimum	Maximum
Latent Class 1 (Fuzzy)	Cluster	N	Mean	Std Dev	Minimum	Maximum
	1	225	0.628	0.0259	0.5554	0.6919
	2	225	0.372	0.0259	0.3081	0.4446
Latent Class 2 (Distinct)	Cluster	N	Mean	Std Dev	Minimum	Maximum
	1	225	0.620	0.0247	0.5483	0.6844
	2	225	0.380	0.0247	0.3156	0.4517
Ward's	Cluster	N	Mean	Std Dev	Minimum	Maximum
	1	225	0.642	0.0567	0.5314	0.8469
	2	225	0.358	0.0567	0.1531	0.4686
Single Linkage	Cluster	N	Mean	Std Dev	Minimum	Maximum
	1	225	0.997	0.0019	0.9871	0.9982
	2	225	0.003	0.0019	0.0018	0.0129
Complete Linkage	Cluster	N	Mean	Std Dev	Minimum	Maximum
	1	225	0.797	0.1023	0.2030	0.9723
	2	225	0.203	0.1023	0.0277	0.7970
Average Linkage	Cluster	N	Mean	Std Dev	Minimum	Maximum
	1	225	0.950	0.0573	0.7011	0.9982
	2	225	0.050	0.0573	0.0018	0.2989

The Distinct Latent Class method had a population proportion of .625 for Cluster 1, and a resampling average of .620, an even smaller offset of .005. The standard deviation of .0247 is very close to the theoretical sampling standard deviation of .0208. The two hundred twenty five replications varied from .548 to .684, a range of .136.

Ward's traditional method had a population proportion of .640 for Cluster 1, and a resampling average of .642, an extremely small offset of .002. However, the standard deviation of .0567 is more than twice the theoretical sampling standard deviation of .0209. The two hundred twenty five replications varied from .531 to .847, a much higher range of .316.

The single linkage traditional method consistently produces one very large cluster, and another extremely small cluster. While the solution is very consistent, it is not very informative. This method was dropped from further consideration.

The complete linkage method had a population proportion of .882 for Cluster 1, and a resampling average of .797, a relatively large offset of .085. The standard deviation of .1023 is almost ten times the theoretical sampling standard deviation of .0139. The two hundred twenty five replications varied from .203 to .972, a higher range of .769.

The average linkage method had a population proportion of .926 for Cluster 1, and a resampling average of .950, an offset of .024. The standard deviation of .0573 is five times the theoretical sampling standard deviation of .0112. The two hundred twenty five replications varied from .701 to .998, a range of .297.

The offset and range for each method is summarized in Table 6. In addition, the relative variance is computed, that is, the variance with resampling divided by the theoretical variance. Note that the single linkage method has been omitted from this table. Of the traditional methods, Ward's method has the smallest offset, relative variance, and range. Although the latent class methods have a somewhat higher offset, their relative variance and range is much lower than Ward's method.

Table 6: Summary

Method	Offset	Relative Variance	Range
Latent Class 1	0.008	1.5	0.137
Latent Class 2	0.005	1.4	0.136
Ward's Linkage	0.002	7.4	0.316
Complete Linkage	0.085	54.2	0.769
Average Linkage	0.024	26.2	0.297

12. SUMMARY AND RECOMMENDATIONS

The original clients of this study hypothesized a priori that the less price sensitive group of customers would be less than 50% of the population, and would probably fall between 20% and 40%. The two latent class methods, Ward's method, and the complete linkage method fall within the bounds of the client's intuition. The original analysis by the author had computed the various classical methods on the complete data set, and reported the results of Ward's method to the client. Although the client was particularly satisfied with the solution, this author had been left with many reservations. How stable was this solution? How would the latent class methods being advocated by some market research firms perform? Would they be as stable as the traditional methods?

The results of the study indicate that the latent class solution is particularly similar to the traditional approach using Ward's clustering algorithm. However, the study also shows that the latent class solution, upon resampling, yields more consistent results in terms of lower variability and a smaller range of results. Although the latent class solution requires relatively specialized custom programming for each problem, this paper would indicate that the results are worth the extra effort.

13. REFERENCE

DeSarbo, W. S., Wedel, M., Vriens, M., and Ramaswamy, V. (1992), "Latent Class Metric Conjoint Analysis", *Marketing Letters*, 1992, 273-288.