TWO MONTE CARLO STUDIES FOR LATENT CLASS SEGMENTATION MODELS

Liberato Camilleri
Department of Statistics and Operations Research
University of Malta
Msida (MSD 06)
Malta
E-mail: liberato.camilleri@um.edu.mt

KEYWORDS
Information Criteria, Proportional Odds Model, Latent Class Model, Segmentation, EM algorithm

ABSTRACT

Model assessment and comparison are essential aspects of statistical inference. The likelihood ratio test is one of the main instruments for model selection; however, this is not appropriate when the model under consideration contains random effects. In this paper, we present two simulation studies for latent class segmentation models. The first Monte Carlo study compares the performance of seven Information Criteria in predicting the correct number of segments. The second study investigates factors that have an effect on segment membership and parameter recovery and affect computational effort.

1. INTRODUCTION – A GENERAL MODEL

Latent class models stands out as one of the major breakthroughs in market segmentation as they overcome the limitations of aggregate analysis and a-priori segmentation. In this approach, the segments and the model parameters within these segments are estimated simultaneously. Latent class methodology for market segmentation, suggested by (Green 2000) proposes the Proportional Odds model as a proper statistical model for ordinal data.

$$P\left(y_{jn} = r \middle| \boldsymbol{\alpha}, \boldsymbol{\beta}\right) = F\left(\alpha_r + \mathbf{x}_j^{'}\boldsymbol{\beta}\right) - F\left(\alpha_{r-1} + \mathbf{x}_j^{'}\boldsymbol{\beta}\right) \qquad (1)$$

In this model y_{jn} is a rating response elicited by the n^{th} respondent for the j^{th} item; $\boldsymbol{\alpha}$ is a vector of threshold parameters; $\boldsymbol{\beta}$ is a vector of regression parameters and $\mathbf{x}_j$ are item or individual covariates. The choices of $F(.)$ considered are the Logistic, Normal and Extreme Value distributions respectively leading to the Logit, Probit and Complementary Log-Log links. For the segmentation model, the Proportional Odds model is extended by considering a Latent Class model with *K* segments.

$$P\left(y_{jn} = r \middle| \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\pi}\right) = \sum_{k=1}^{K} \pi_k . P\left(y_{jn} = r \middle| \boldsymbol{\alpha}, \boldsymbol{\beta}_k\right) \qquad (2)$$

where π_k is the proportion of respondents that are assigned to the k^{th} segment. The log-likelihood function is maximized through the EM algorithm. The merit of this model is that it allows for a probabilistic classification of respondents into segments and simultaneous estimation of a generalized linear regression model within each segment. When applying the above model to real data, the actual number of segments, *K*, is unknown and has to be specified. Unfortunately, the standard likelihood ratio statistic that tests between a *K*-segment model and a (*K*+1)-segment model does not have an asymptotic chi-square distribution and so it is not adequate to identify the appropriate number of segments in latent class models. Several information criteria have been proposed to identify this optimal number of segments.

2. INFORMATION CRITERIA TO IDENTIFY THE NUMBER OF SEGMENTS

Several Information Criteria have been proposed to compare Latent class models with different number of components (segments). These criteria stipulate different penalty terms to measure the complexity of the model. Most of the information criteria that are proposed are based on the bias-corrected log-likelihood given by:

$$C = -2\log L\left(\boldsymbol{\Psi}\right) + dc \qquad (3)$$

where *d* is the number of estimated parameters and *c* is a penalty constant. The second term, which is the penalty term, measures the complexity of the model. For instance the well-known Akaike information criterion (AIC), proposed by (Akaike 1974), arises when $c = 2$. The major problem with the use of this criterion is that it relies on the same asymptotic properties as the likelihood ratio test. Many authors have observed that AIC tend to overestimate the correct number of segments. The Modified Akaike information criterion (MAIC) arises when $c = 3$ and it penalizes complex models more heavily than AIC. Another criterion that penalizes the log-likelihood more heavily is the Bayesian information criterion (BIC), proposed by (Schwarz 1978). For this criterion $c = \log(N)$, where N is the sample size. The penalty term of the BIC criterion depends on the sample size and favours models that are more parsimonious. For $N > 8$, BIC penalizes complex models more heavily than AIC and MAIC. BIC reduces the tendency of AIC and MAIC to fit too many segments.

The above criteria account for over-parameterization as more segments are derived. However, one must ensure that the segments are sufficiently separated for a particular solution. To examine the centroid separation between the segments, (Ramaswamy and Cohen 2000) use an entropy statistic to investigate the degree of separation in the estimated posterior probabilities.

$$E_s = 1 + \frac{EN(\hat{p}_{nk})}{N \log K} \quad (4)$$

where $\hat{p}_{nk}$ is the posterior probability that the n^{th} subject belongs to the k^{th} segment and $EN(\hat{p}_{nk})$ is the entropy given by $EN(\hat{p}_{nk}) = -\sum_{k=1}^{K}\sum_{n=1}^{N} \hat{p}_{nk} \log(\hat{p}_{nk})$. $EN(\hat{p}_{nk})$ is the penalty term and penalizes models whose segments are poorly separated. When the segment centroids of these latent class models are well separated, $EN(\hat{p}_{nk})$ will be close to its minimum value of zero. If the segment centroids are not sufficiently separated for the number of segments specified then $EN(\hat{p}_{nk})$ will have a large value because the posterior probabilities for each observation are approximately equal. E_s is a relative measure bounded between 0 and 1. A value close to 1 indicates that the centroids of the derived segments are well separated. A value close to 0 indicates poor separation.

Another criterion that uses $EN(\hat{p}_{nk})$ to penalize a model for its complexity is the Classification Likelihood criterion (CLC) proposed by (Biernacki and Govaert 1997). This criterion minimizes

$$CLC = -2\log L(\mathbf{\Psi}) + 2EN(\hat{p}_{nk}) \quad (5)$$

and the penalty on the log-likelihood depends on how well separated the fitted segments are. This criterion works best when the probabilities of segment membership happen to be similar. However, it tends to overestimate the correct number of segments when no restriction is placed on these probabilities.

Another criterion that uses the normalized form of $EN(\hat{p}_{nk})$ for choosing the number of segments is the Normalized Entropy Criterion (NEC) proposed by (Celeux and Soromenho 1996). This normalized form is given by:

$$NEC = \frac{EN(\hat{p}_{nk})}{\log L(\mathbf{\Psi}) - \log L(\mathbf{\Psi}^*)} \quad (6)$$

where $\log L(\mathbf{\Psi}^*)$ is the log-likelihood in the case of a single segment $(K = 1)$. This criterion has a shortcoming because $EN(\hat{p}_{nk}) = 0$ for $K = 1$ and so it is unable to decide between $K = 1$ and a value of K greater than one. (Biernacki, Celeux and Govaert 1999) proposed a modification to overcome this limitation. The modified criterion defines $NEC = 1$ for $K = 1$ and then chooses the number of segments to minimize NEC.

Another criterion, proposed by (Biernacki, Celeux and Govaert 1999), is the Integrated Classification Likelihood (ICL), which assumes that $N\pi_k$ are sufficiently large values. This criterion is chosen in an attempt to overcome the shortcomings of BIC and CLC. It minimizes

$$ICL = -2\log L(\mathbf{\Psi}) + 2EN(\hat{p}_{nk}) + d\log(N) \quad (7)$$

3. STUDY DESIGN TO ASSESS INFORMATION CRITERIA PERFORMANCE

In order to assess the performance of the proposed criteria in identifying the correct number of segments, synthetic data sets were generated using a GLIM algorithm. The simulation was devised to mimic the application of (Camilleri and Green 2004) in which four car brands; four price values and two door features were generated to define the item attributes. In the application, a full profile approach was employed in which 32 items (cards) were generated where each card had a unique item attribute combination. This guaranteed a full factorial design. In the simulation, the item attribute values and the number of hypothetical respondents (N) were set the same as in the application. Two sets of uniformly distributed pseudo-random real values in the range [0,1] were used to generate the age and gender of each hypothetical subject. Pseudo random values less than 0.5 in the first set corresponded to male subjects. By transforming the pseudo random values in the second set using a linear relationship, the ages of the hypothetical subjects were generated to vary from 15 to 75 years.

To allocate the N hypothetical subjects into K segments, the proportions π_k were set the same as in the application. The cumulative probabilities, $q_0, q_1, ..., q_K$, were computed such that $q_k = \sum_{i=1}^{k}\pi_i$, where $q_0 = 0$ and $q_K = 1$. A set of uniformly distributed pseudo-random real values was then generated in the range [0, 1] to allocate the hypothetical subjects to one of the K segments. Subjects whose corresponding pseudo-random values were in the range $[q_{k-1}, q_k]$ were allocated to the

k^{th} segment. This classification gave each subject a random segment allocation.

To simulate the subjects' utility responses, the utility model (linear predictor) and the parameter values for the *K* segments were set the same as in the application. The utility model included both main effects and interaction terms of the item and individual covariates. 32 synthetic data values or utility values were generated for each hypothetical subject by substituting the parameter values and the values of the item and individual covariates in the utility model.

Error terms ε_i were added to these utility values to have either a logistic or a normal or an extreme value distribution. These error terms were generated by transforming pseudo-random real values u_i in the range [0,1] from a uniform distribution. If ε_i has a logistic distribution then $\varepsilon_i = \ln\left[u_i/(1-u_i)\right]$; $\varepsilon_i = \Phi^{-1}(u_i)$ if ε_i have a normal distribution and $\varepsilon_i = \ln\left[-\ln(1-u_i)\right]$ if ε_i have an Extreme value distribution. A set of six specified cut-point values α_r was used to convert these modified utility values to rates ranging from 1 to 7. Items (cards) whose modified worth values were in the range (α_{r-1}, α_r) were rated in the r^{th} worth category. This classification gives the rating responses of each hypothetical subject a random category allocation.

4. RESULTS OF THE FIRST STUDY

An empirical comparison was carried out to determine which of the above criteria best select the correct number of segments in a Latent Class model. The study compared the performance of the more recently suggested criteria such as CLC, NEC and ICL with classical procedures such as AIC, MAIC and BIC. By assuming a Logistic distribution, fifteen data sets were generated using the same utility model, design matrix and parameter values. These data sets were simulated using $N = 310$ and $K = 4$. Each simulated data set was re-fitted four times varying the number of segments from three to six clusters. The log-likelihood and the entropy were recorded to determine the number of segments that minimize the specified criteria. Solutions that were spurious were eliminated and a different random start was considered to initialize the EM algorithm. The numbers of parameters for the latent class model with 3, 4, 5 and 6 segments were respectively *d* = 58, 76, 94 and 112. These include the parameters of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\pi}$.

It is evident from tables 1 and 3 that the reduction in the log-likelihood is significantly larger when fitting 3 and 4 segments when compared to fitting 4, 5 and 6 segments. This implies that the reduction in the log-likelihood tends to become smaller when the number of fitted segments exceeds the number of true segments.

	K=3	**K=4**	**K=5**	**K=6**
Set	$-2\log L(\hat{\boldsymbol{\Psi}})$	$-2\log L(\hat{\boldsymbol{\Psi}})$	$-2\log L(\hat{\boldsymbol{\Psi}})$	$-2\log L(\hat{\boldsymbol{\Psi}})$
1	15059.3	14083.8	14062.4	13933.5
2	14987.7	14215.6	14099.2	14090.3
3	15018.0	14106.8	14039.9	13889.7
4	15107.5	14195.2	14126.9	14047.3
5	15006.8	14177.2	14107.8	14034.9
6	15130.1	14191.1	14141.6	14115.3
7	14972.7	14143.8	14096.5	14025.8
8	15077.6	14148.0	14082.7	14043.9
9	15160.7	14096.1	14062.8	13975.8
10	15032.9	14059.5	13949.6	13822.5
11	14857.4	14018.3	13943.7	13825.2
12	14984.5	14170.1	14125.6	14054.2
13	14931.4	14186.0	14155.4	14087.4
14	14995.5	14154.7	14098.2	14021.8
15	15206.0	14363.9	14296.6	14184.3

Table 1: Deviances for 3, 4, 5 and 6 segments

	K=3	**K=4**	**K=5**	**K=6**
Set	**Entropy**	**Entropy**	**Entropy**	**Entropy**
1	45.54	9.036	16.23	25.89
2	17.89	10.43	19.93	13.40
3	42.88	8.590	18.93	28.48
4	44.91	10.40	26.44	44.34
5	48.86	12.15	15.82	40.77
6	81.86	11.72	23.59	34.77
7	12.21	10.04	20.50	18.60
8	71.10	10.01	10.76	33.83
9	53.89	10.75	18.65	24.11
10	73.70	15.25	16.70	26.16
11	43.72	10.06	21.99	23.57
12	74.00	11.15	9.910	20.94
13	76.76	15.98	16.36	35.73
14	70.83	10.99	26.35	25.13
15	8.210	9.998	20.45	43.42

Table 2: Entropies for 3, 4, 5 and 6 segments

		$-2\log L(\hat{\boldsymbol{\Psi}})$	Entropy
***K*=3**	Mean	15035.2	51.09
	Standard Deviation	90.77	24.04
	Minimum	14857.4	8.21
***K*=4**	Mean	14154.0	11.10
	Standard Deviation	80.51	2.049
	Minimum	14018.3	8.59
***K*=5**	Mean	14092.6	18.84
	Standard Deviation	83.60	4.818
	Minimum	13943.7	9.91
***K*=6**	Mean	14016.8	29.02
	Standard Deviation	98.36	9.202
	Minimum	13822.5	13.4

Table 3: Descriptive statistics for deviances and entropies

Table 3 displays another interesting result. The mean and standard deviation of the entropy are smallest when 4 segments are fitted. These two measures increase when the number of fitted segments exceeds the number of true segments. The addition of extra segments increases the mean entropy because the centroid separation between the segments is reduced. The number of hypothetical subjects in these extra segments also affects the size of the entropy.

Tables 4 and 5 show segment membership recovery of the 310 hypothetical subjects for the 4^{th} and 7^{th} data sets. The entropies for these two data sets were respectively 44.34 and 18.60. This implies that the entropy increases as the number of hypothetical subjects in these extra segments increases. When the number of fitted segments exceeds the number of true segments there are two possible outcomes. If four of the fitted segments include a large proportion of the hypothetical subjects such that the other two segments have small frequencies then the segments are more likely to be well separated and the posterior probabilities tend to be close to either 0 or 1. This yields a small entropy value. However, if one or more of the larger segments split, locations of the new clusters are relatively close. Moreover, the proportion of correctly classified hypothetical subjects decreases. So the posterior probabilities are more likely to be distant from either 0 or 1, yielding a larger entropy value. This explains why the dispersion of the entropies increases when too many segments are fitted.

		Fitted Segments					
		1	2	3	4	5	6
True Segments	1	52	0	0	0	36	0
	2	0	67	0	0	0	0
	3	1	0	45	0	1	31
	4	0	0	0	77	0	0

Table 4: Segment allocations for the 4^{th} data set

		Fitted Segments					
		1	2	3	4	5	6
True Segments	1	97	0	0	0	0	0
	2	0	58	0	0	0	3
	3	1	0	56	0	15	0
	4	0	0	0	74	0	6

Table 5: Segment allocations for the 7^{th} data set

Table 6 shows the result of the empirical comparison between 7 different criteria to determine which one best selects the correct number of segments in a Latent Class model. It is evident that the modern procedures NEC and ICL outperform the classical procedures AIC and MAIC. AIC and MAIC have a tendency to fit too many segments, whereas BIC penalize complex models more heavily than AIC and MAIC. The penalty term of most criteria depends on one or more of the following features; number of estimated parameters, number of subjects and entropy. Criteria that combine two or more of these features in the penalty term are superior to those that contain only one feature. The penalty term of AIC and MAIC depends solely on number of estimated parameters and the penalty term of CLC depends entirely on the entropy. All three are inferior to the other criteria in recovering the true number of segments.

K	AIC	MAIC	BIC	CLC	ICL	NEC	E_s
3	0	0	0	0	0	0	0
4	0	1	12	0	14	14	11
5	2	2	1	1	1	0	4
6	13	12	2	14	0	1	0

Table 6: Criteria Performance (Logistic distribution)

A further task was included in the study to investigate how the choice of the distribution function affects the performance of the Information Criteria in selecting the optimal number of clusters and how it affects segment membership recovery. To carry out this task, a further thirty data sets were generated using the same utility model, design matrix and parameter values as in the previous assignment. A Normal distribution was assumed to generate the first fifteen data sets and an Extreme value distribution was assumed to generate the rest. These data sets were simulated using $N = 310$ and $K = 4$. Each simulated data set was again re-fitted four times varying the number of segments from three to six clusters. The log-likelihood and the entropy were recorded to determine the number of segments that minimize the specified criteria.

Correct segment allocation deteriorates slightly when an Extreme value distribution is assumed; however, segment-membership recovery improves when using a Normal distribution. Tables 7 and 8 show the results of the empirical comparison of the Information Criteria when a Normal or an Extreme Value distribution was assumed. Both tables display that the procedures ICL, NEC and BIC outperform the procedures AIC, MAIC and CLC. This implies that the performance of the Information Criteria in selecting the correct number of segments is not affected much by the choice of the distribution.

K	AIC	MAIC	BIC	CLC	ICL	NEC	E
3	0	0	0	0	0	0	0
4	0	0	14	1	15	15	12
5	4	5	1	2	0	0	3
6	11	10	0	12	0	0	0

Table 7: Criteria Performance (Normal distribution)

K	AIC	MAIC	BIC	CLC	ICL	NEC	E
3	0	0	0	0	0	0	0
4	0	0	10	0	13	12	10
5	0	0	4	0	2	1	4
6	15	15	1	15	0	2	1

Table 8: Criteria Performance (Extreme value distribution)

5. FACTORS AFFECTING THE PERFORMANCE OF LATENT CLASS MODELS

A further task was to examine the performance of latent class models by modifying a number of factors. Three of the factors that are highlighted in literature (Vriens, Wedel and Wilms 1996; Wedel and DeSarbo 1995) as having potential effect on model performance include:

- Number of simulated respondents
- Number of segments
- Distribution of the dependent variable

The above three factors reflect a variation in conditions in many practical applications and which are expected to affect the performance of the model fit. The following six measures are normally used to assess computational effort, parameter recovery, predictive power, goodness of fit and segment membership recovery.

- The percentage of variance, R^2 accounted for by the latent class model is a measure of the goodness of fit.

- The number of iterations required for convergence is a measure of the computational effort.

- The root-mean-squared error between the true and estimated parameters and the root-mean-squared error between the true and estimated segment membership probabilities are measures of parameter recovery.

$$RMS\left(\hat{\boldsymbol{\beta}}\right)=\left[\sum_{p=1}^{P}\frac{\left(\beta_p-\hat{\beta}_p\right)^2}{P}\right]^{\frac{1}{2}} \tag{8}$$

  $\hat{\beta}_p$ and β_p are respectively the estimated and true parameters; whereas P is the number of parameters.

$$RMS\left(\hat{\boldsymbol{\pi}}\right)=\left[\sum_{k=1}^{K}\frac{\left(\pi_k-\hat{\pi}_k\right)^2}{K}\right]^{\frac{1}{2}} \tag{9}$$

  $\hat{\pi}_k$ and π_k are respectively the estimated and true segment membership probabilities; whereas K is the number of segments.

- The root-mean-squared-error between the true and predicted responses is a measure of the predictive power.

$$RMS\left(y\right)=\left[\sum_{n=1}^{N}\sum_{j=1}^{J}\frac{\left(y_{nj}-y_{nj}\right)^2}{N.J}\right]^{\frac{1}{2}} \tag{10}$$

  y_{nj} and y_{nj} are respectively the estimated and true responses; whereas N and J are respectively the number of simulated respondents and the number of items (cards) assessed by each subject.

- The percentage number of subjects that are correctly classified into their true segments is a measure of segment membership recovery. A subject is assigned to the segment with highest posterior probability.

6. STUDY DESIGN TO ASSESS THE FACTORS THAT AFFECT MODEL PERFORMANCE

In order to assess the factors that affect the performance of latent class models, synthetic data sets were generated using a GLIM algorithm. The simulation was devised to mimic the application of (Camilleri and Green 2004). The design and the utility model (linear predictor) were set the same as in the application. The age and gender of the N hypothetical subjects and their segment allocation were generated using a similar procedure described in the first simulation study. To simulate subjects' rating responses, a set of parameters was specified such that all the main effects and interaction terms in each segment were assigned a parameter. 32 synthetic response values were generated for each hypothetical subject by substituting the parameter values and the values of the item and individual covariates in the utility model.

The number of simulated respondents was varied at two levels: 200 and 310. These levels represent a reasonable range of sample sizes that are reported in several segmentation applications (Wittink, Vriens and Burhenne 1994; Wedel and Steenkamp 1991). Standard theory on statistical inference suggests that a greater number of simulated respondents improve the precision of the estimated segment-level parameters.

The number of segments was also varied at two levels. A two-segment and a four-segment condition were used because these represent the range of segments commonly found in segmentation applications (Wedel and Steenkamp 1989; DeSarbo, Oliver and Rangaswamy 1989). It is expected that a greater number of segments deteriorate the precision of the estimated segment-level coefficients as a greater number of model parameters have to be estimated.

The Proportional Odds model can accommodate three possible distribution functions. The Logistic, Normal and Extreme value distributions, which respectively lead to logit, probit and complementary log-log link functions, were all considered.

A problem associated with the application of the EM algorithm to latent class models is its convergence to local maxima. It is caused by the likelihood being multimodal, so that the algorithm becomes sensitive to the starting values used. The problem of convergence to local optima becomes more conspicuous when the component densities are not well separated and when the number of estimated parameters is large. This will lead to a relatively weak update in the E-step (Wedel and Kamakura 2000). To overcome this problem five starting values were considered for each combination of the factor levels defined above. These were selected from a wide range of seed numbers. Another problem with the EM algorithm is that the fitted segments are very often a swapped version of the true segments. To overcome this problem, the parameters were chosen to contrast considerably between segments. The purpose was to simplify the identification of the correct correspondence between the fitted segments and the true segments.

7. RESULTS OF THE SECOND STUDY

Five data sets were generated for each factor level combination according to the type of distribution, number of subjects and number of segments. Each simulated data set was re-fitted using a latent class model. Solutions that were considered spurious were eliminated and a different random start was considered to initialize the EM algorithm. The statistics $RMS(\hat{\boldsymbol{\beta}})$ $RMS(\hat{\boldsymbol{\pi}})$ and $RMS(\hat{y})$ were computed after permuting the parameters and predicted responses to match estimated and true segments optimally. All the six measures were averaged over these five data sets.

Table 9 exhibits some differences in the R^2 measures between the types of distributions. The goodness of fit is improved when the choice of the distribution is Normal or Logistic. The value of R^2 increases with an increasing number of segments and a decreasing number of subjects.

Distribution	Number of respondents	Number of segments	R^2
Logistic	200	2	0.8516
Normal			0.8612
Extreme			0.8394
Logistic	310		0.8455
Normal			0.8563
Extreme			0.8222
Logistic	200	4	0.9217
Normal			0.9366
Extreme			0.8876
Logistic	310		0.9158
Normal			0.9193
Extreme			0.8685

Table 9: Measures of goodness of fit

Table 10 demonstrates that the number of segments mostly affects computational effort. An increase in the number of segments increases the number of iterations required. The number of simulated respondents and the choice of the error distribution have negligible effect on computational effort.

Distribution	Number of respondents	Number of segments	Number of iterations
Logistic	200	2	29.2
Normal			30.4
Extreme			31.6
Logistic	310		32.2
Normal			28.8
Extreme			29.4
Logistic	200	4	36.6
Normal			37.8
Extreme			35.8
Logistic	310		38.2
Normal			37.6
Extreme			38.2

Table 10: Measures of computational effort

Table 11 exhibits that the number of segments and the choice of distribution affect the percentage of correctly classified subjects; however, the number of subjects has negligible effect on segment-membership recovery. Segment-membership recovery deteriorates slightly with an increase in the number of segments and this deterioration worsens when an Extreme value distribution is used. When hypothetical subjects are allocated to segments, a mismatch in a four-segment solution is more likely to occur than in a two-segment solution. Segment-membership is recovered best when using a Normal or a Logistic distribution.

Distribution	Number of respondents	Number of segments	Segment recovery
Logistic	200	2	98.8%
Normal			99.0%
Extreme			95.1%
Logistic	310		99.3%
Normal			99.4%
Extreme			95.2%
Logistic	200	4	98.5%
Normal			98.5%
Extreme			94.3%
Logistic	310		98.7%
Normal			98.9%
Extreme			94.0%

Table 11: Measures of segment membership recovery

Table 12 exhibits that the mean $RMS(\hat{\boldsymbol{\pi}})$ is not affected by the choice of distribution used. The mean $RMS(\hat{\boldsymbol{\pi}})$ decreases with an increase in the number of segments and hypothetical subjects. Increasing the number of model parameters and increasing the sample size improve the probabilities of segment membership.

Distribution	Number of respondents	Number of segments	rms $(\hat{\boldsymbol{\pi}})$
Logistic	200	2	0.0429
Normal			0.0461
Extreme			0.0433
Logistic	310		0.0325
Normal			0.0313
Extreme			0.0329
Logistic	200	4	0.0238
Normal			0.0256
Extreme			0.0219
Logistic	310		0.0197
Normal			0.0195
Extreme			0.0199

Table 12: Measures of segment proportion recovery

Table 13 demonstrates that the type of distribution affects parameter recovery. The mean $RMS(\hat{\boldsymbol{\beta}})$ is lowest when the choice of the error distribution is Normal and highest when the Extreme value distribution is used. Parameter recovery improves with an increase in the number of simulated subjects but deteriorates with an increase in the number of segments.

Distribution	Number of respondents	Number of segments	rms $(\hat{\beta})$
Logistic	200	2	0.2215
Normal			0.2109
Extreme			0.2678
Logistic	310		0.1633
Normal			0.1527
Extreme			0.2134
Logistic	200	4	0.2596
Normal			0.2578
Extreme			0.3073
Logistic	310		0.2236
Normal			0.2227
Extreme			0.2511

Table 13: Measures of parameter recovery

Table 14 shows that an increase in the number of subjects and a decrease in the number of segments improve the predictive accuracy. The Extreme-value distribution yields the highest mean value of $RMS(\hat{y})$ implying that the predictive accuracy deteriorates when this distribution is used.

Distribution	Number of respondents	Number of segments	rms $(\hat{y})$
Logistic	200	2	2.1629
Normal			2.1538
Extreme			2.6664
Logistic	310		2.1215
Normal			2.1036
Extreme			2.5319
Logistic	200	4	3.0516
Normal			3.0429
Extreme			3.1219
Logistic	310		3.0217
Normal			3.0424
Extreme			3.1056

Table 14: Measures of predictive power

8 CONCLUSIONS

In the first Monte Carlo study, several Information Criteria were proposed to determine the optimal number of segments in a Latent Class model. These criteria specify different penalty terms to measure the complexity of the model. These penalty terms depend on one or more model features, which include the number of estimated parameters, number of hypothetical subjects and entropy. This study illustrates that information Criteria that combine two or more model features in the penalty term outperform those criteria that include solely one feature. This study also demonstrates that the performance of these Criteria is not affected by the choice of the distribution function.

The second simulation study reveals several appealing results. Goodness of fit improves and computational effort increases with a larger number of segments; however, parameter recovery, segment membership recovery and predictive accuracy improve with a smaller number of segments. These results conform to standard theory on statistical inference. The choice of the error distribution has noticeable effect on model performance. The Normal and Logistic distributions outperform the Extreme value distribution. These two distributions yield better fits and improve predictive power, segment membership recovery and parameter recovery. In general, the effects of the number of simulated subjects on parameter recovery, segment membership recovery, computational effort, predictive accuracy and goodness of fit are smaller than the effects of the number of segments and distribution choice.

ACKNOWLEDGEMENTS

The author gratefully acknowledges Dr Michael Green for his contribution in developing a simulation algorithm that generates synthetic data for Latent Class model fitting.

REFERENCES

Aitkin, M., Anderson, D., Francis, B. and Hinde, J. (1994), *Statistical Modelling in GLIM*, Oxford Science Publications.

Akaike, H. (1974), A New Look at Statistical Model Identification, *IEEE Transactions on Automatic Control*, AC-19, 716-723.

Biernacki, C., Celeux, G. and Govaert, G. (1999), An Improvement of NEC Criterion for assessing the number of Clusters in a Mixture Model, *Pattern Recognition Letters*, 20, 267-272.

Biernacki, C. and Govaert, G. (1997), Using the Classification Likelihood to Choose the Number of Clusters, *Computing Science and Statistics*, 29, 451-457.

Camilleri, L. and Green, M. (2004), Statistical Models for Market Segmentation, *Proceedings of the 19th International Workshop Statistical Modelling, Florence*. 120-124.

Camilleri, L. and Portelli, M. (2007), Segmenting the heterogeneity of tourist preferences using a Latent Class model combined with the EM algorithm, *Proceedings of the 6th APLIMAT International Conference, Bratislava*. 343-356.

Celeux, G. and Soromenho, G. (1996), An Entropy Criterion for assessing the number of clusters in a Mixture model, *Classification Journal*, 13, 195-212.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977), Maximum Likelihood from Incomplete Data via the EM algorithm, *Journal of the Royal Statistical Society*, B, 39, 1-38.

DeSarbo, W.S., Oliver, R.L. and Rangaswamy, A. (1989), A Simulated Annealing Methodology for Clusterwise Linear Regression, *Psychometrika*, 54, 707-736.

Francis, B., Green, M. and Payne, C. (1993), *The GLIM 4 manual*, Oxford Science Publications.

Green, M. (2000), Statistical Models for Conjoint Analysis, *Proceedings of the 15th International Workshop on Statistical Modelling, Bilbao*. 216-222.

Nelder, J.A and Wedderburn, R.W.M. (1972), Generalized Linear Models, *Journal of the Royal Statistical Society*, A, 135, 370-384.

Ogawa, K. (1987), An Approach to Simultaneous Estimation and Segmentation in Conjoint Analysis, *Marketing Science*, 6, 66-81.

Ramaswamy, V. and Cohen, S.H. (2000), Latent Class Models for Conjoint Analysis, in Gustafsson, A., Hermann, A. and Huber, F., ed., *Conjoint Measurement: Methods and Applications*, Springer, 415-446.

Schwarz, G. (1978), Estimating the Dimension of a Model, *The Annals of Statistics*, 6, 461-464.

Vermunt, J.K. (2004), An EM algorithm for the estimation of parametric and non-parametric hierarchical non-linear models, *Statistica Neerlandica*, 58, 220-233.

Vriens, M., Wedel, M. and Wilms, T. (1996), Conjoint Segmentation Methods A Monte Carlo Comparison, *Journal of Marketing Research*, 23, 73-85.

Wedel, M. and DeSarbo, W.S. (1995), A Mixture Likelihood Approach for Generalized Linear Models, *Journal of Classification*, 12, 1-35.

Wedel, M. and Kamakura, W.A. (2000), *Market Segmentation: Conceptual and Methodological Foundations*, Kluwer Academic Publishers.

Wedel, M. and Kistemaker, C. (1989), Consumer Benefit Segmentation using Clusterwise Linear Regression, *International Journal of Research Market*, 6, 45-49.

Wedel, M. and Steenkamp, J.B. (1989), Fuzzy Clusterwise Regression Approach to Benefit Segmentation, *International Journal of Research in Marketing*, 6, 241-258.

Wedel, M. and Steenkamp, J.B. (1991), A Clusterwise Regression Method for Simultaneous Fuzzy Market Structuring and Benefit Segmentation, *Journal of Market Research*, 28, 385-396.

Wittink, D.R., Vriens, M. and Burhenne, W. (1994), Commercial Use of Conjoint Analysis in Europe, Results and Critical Reflections, *International Journal of Research in Marketing*, 11, 41-52.

AUTHOR BIOGRAPHY

LIBERATO CAMILLERI studied Mathematics and Statistics at the University of Malta. He received his PhD degree in Applied Statistics in 2005 from Lancaster University. His research specialization areas are related to statistical models, which include Generalized Linear models, Latent Class models, Multi-Level models and Random Coefficient models. He is presently a lecturer in the Statistics department at the University of Malta.