View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by OAR@UM

Latent Class Mixture models for analyzing rating responses

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

KEYWORDS
Proportional Odds Model, Latent Class Mixture Model, EM algorithm, Maximum Likelihood Estimation

ABSTRACT

Latent class methodology has been used extensively in market research. In this approach, segment membership and parameter estimates for each derived segment are estimated simultaneously. A popular approach for fitting latent class models to rating responses is to assume mixtures of multivariate conditional normal distributions. An alternative approach is to assume a Proportional Odds model. These two approaches are compared empirically in a Monte Carlo study, assessing segment membership and parameter recovery, goodness of fit and predictive accuracy.

INTRODUCTION

Latent class segmentation models are used extensively in various fields of application to identify latent segments that can explain unobserved heterogeneity in the data. Traditionally, segmentation procedures were carried out using a two-stage approach in which estimation and clustering were conducted consecutively. In the first step, individual-level parameter estimates were derived from a Normal regression model. In the second step, individuals were clustered on the basis of similarity of the estimated parameters using a clustering algorithm. Typically, latent class regression analysis comprises the following three simultaneous steps – identify hidden segments; classify each individual in an appropriate class and estimate a regression model for each segment. Latent class models have been utilized for various types of responses, mainly rating responses. Multivariate normal latent class models have been applied by DeSarbo, et al. (1992) to analyze rating responses in a conjoint study that examines the design of a remote automobile entry device. Helsen, et al. (1993) used a similar model to classify countries into homogeneous groups having similar patterns of diffusion of durable goods. Ramaswamy, et al. (1993) applied the mixture regression model to cross-sectional time-series data. Camilleri and Green (2003) combined a latent class model with the proportional odds model to analyze rating responses in a study related to car preferences.

LATENT CLASS MODELS FOR RATING DATA

One approach of utilizing latent class models to rating responses is to assume that respondents perceive scale spacing so that preferences are used as metric data. A latent class methodology can then be employed by using mixtures of multivariate conditional normal distributions combined with the EM algorithm to estimate parameters of these mixtures. The conditional multivariate density function of the response vectors $\mathbf{y}_n = \left(y_{nj}\right)$ for $j = 1,\ldots,J$ replications, given that the n^{th} respondent belongs to the k^{th} segment is:

$$f_{n|k}\left(\mathbf{y}_n;\boldsymbol{\beta}_k\right) = \left(2\pi\right)^{-J/2}\left|\boldsymbol{\Sigma}_k\right|^{-1/2}\exp\left[-\frac{1}{2}\left(\mathbf{y}_n - \mathbf{X}\boldsymbol{\beta}_k\right)'\boldsymbol{\Sigma}_k^{-1}\left(\mathbf{y}_n - \mathbf{X}\boldsymbol{\beta}_k\right)\right]$$

where $\boldsymbol{\Sigma}_k$ is the variance-covariance matrix of $\mathbf{y}_n$ given segment k. The unconditional density function is:

$$f_n\left(\mathbf{y}_n;\boldsymbol{\pi},\boldsymbol{\beta}\right) = \sum_1^K \pi_k . f_{n|k}\left(\mathbf{y}_n;\boldsymbol{\beta}_k\right)$$

The likelihood approach is very often used for estimation of finite mixtures because maximum likelihood estimates have been found to be superior to other methods. The log likelihood function can be formulated as:

$$\ln L\left(\boldsymbol{\pi},\boldsymbol{\beta}\right) = \ln\prod_{n=1}^{N} f_n\left(\mathbf{y}_n;\boldsymbol{\pi},\boldsymbol{\beta}\right) = \sum_{n=1}^{N}\ln\sum_{k=1}^{K}\pi_k . f_{n|k}\left(\mathbf{y}_n\left|\boldsymbol{\beta}_k\right.\right)$$

The derivatives of the expected log-likelihood function $E\left[\ln L\left(\boldsymbol{\pi},\boldsymbol{\beta}\right)\right]$ with respect to the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\pi}$ are not straightforward. An effective procedure that fits a latent class model with K segments is to maximize the expected complete log-likelihood function using the EM algorithm. The EM algorithm augments the observed data by introducing unobserved 0-1 indicators λ_{nk}, where λ_{nk} indicates whether the n^{th} respondent belongs to the k^{th} segment. Given the matrix $\boldsymbol{\Lambda} = \left(\lambda_{nk}\right)$ of unobserved data, the complete log-likelihood function is:

$$\ln L\left(\boldsymbol{\pi},\boldsymbol{\beta}\left|\boldsymbol{\Lambda}\right.\right) = \sum_{n=1}^{N}\sum_{k=1}^{K}\lambda_{nk}.\ln f_{n|k}\left(\mathbf{y}_n\left|\boldsymbol{\beta}_k\right.\right) + \sum_{n=1}^{N}\sum_{k=1}^{K}\lambda_{nk}.\ln\left(\pi_k\right)$$

$\ln L(\boldsymbol{\pi},\boldsymbol{\beta}|\boldsymbol{\Lambda})$ has a simpler form than $\ln L(\boldsymbol{\pi},\boldsymbol{\beta})$ and is easy to differentiate. Once the parameter estimates for $\boldsymbol{\beta}_k$ and π_k are obtained, an estimate for the posterior probability $\hat{p}_{nk} = E(\lambda_{nk})$ can be calculated using Bayes' theorem.

$$\hat{p}_{nk} = E(\lambda_{nk}) = \frac{\hat{\pi}_k \cdot f_{n|k}\left(\mathbf{y}_n \middle| \hat{\boldsymbol{\beta}}_k\right)}{\sum_1^K \hat{\pi}_k \cdot f_{n|k}\left(\mathbf{y}_n \middle| \hat{\boldsymbol{\beta}}_k\right)} \quad \text{where } \sum_{i=1}^{K} \hat{p}_{nk} = 1$$

Each iteration is composed of an E-step and an M-step. In the E-step, the expected log-likelihood function is calculated by replacing $E(\lambda_{nk})$ by $\hat{p}_{nk}$.

$$E\left[\ln L(\boldsymbol{\pi},\boldsymbol{\beta}|\boldsymbol{\Lambda})\right] = \sum_{n=1}^{N}\sum_{k=1}^{K} \hat{p}_{nk}.\ln f_{n|k}\left(\mathbf{y}_n|\boldsymbol{\beta}_k\right) + \sum_{n=1}^{N}\sum_{k=1}^{K} \hat{p}_{nk}.\ln(\pi_k)$$

In the M-step the two terms of $E\left[\ln L(\boldsymbol{\pi},\boldsymbol{\beta}|\boldsymbol{\Lambda})\right]$ are maximized separately with respect to the parameters. Maximizing the first term with respect to $\boldsymbol{\beta}_k$ leads to independently solving each of the K expressions

$$\sum_{n=1}^{N} \hat{p}_{nk} \cdot \frac{\partial}{\partial \boldsymbol{\beta}_k} \ln f_{n|k}\left(\mathbf{y}_n|\boldsymbol{\beta}_k\right) \quad \text{for} \quad k = 1,\dots,K$$

Maximizing the second term with respect to π_k, subject to the constraint $\sum_1^K \pi_k = 1$, yields

$$\hat{\pi}_k = \frac{1}{N}\sum_{n=1}^{N} \hat{p}_{nk} \quad \text{for} \quad k = 1,\dots,K$$

The iterative procedure is initiated by setting random values to $\hat{p}_{nk}$. The algorithm then alternately updates the parameters $\hat{\boldsymbol{\beta}}_k, \hat{\pi}_k$ and the posterior probabilities $\hat{p}_{nk}$ until the process converges. Individuals are then assigned to the segment with the highest posterior probability $\hat{p}_{nk}$.

An alternative method to use ordered response categories is by forming models of cumulative probabilities. The Proportional Odds model described by McCullagh (1980) is a cumulative model that preserves the discrete ordinal nature of the rating responses.

Let y_j be the rating response for the j^{th} item using an R-point Likert scale and let $P(y_j \le r)$ be the r^{th} cumulative probability of this item. Cumulative probabilities reflect the ordering since

$$P(y_j \le 1) \le P(y_j \le 2) \le \dots \le P(y_j \le R) = 1$$

To include the effects of explanatory variables we use the model:

$$P(y_j \le r) = F\left(\alpha_r + \mathbf{x}_j^{'}\boldsymbol{\beta}\right) \quad \text{for } r = 1,\dots,R-1$$

$\boldsymbol{\alpha} = (\alpha_1,\dots,\alpha_{R-1})$ is a vector of cutpoint parameters such that $\alpha_1 \le \alpha_2 \le \dots \le \alpha_{R-1}$, $\alpha_0 = -\infty$ and $\alpha_R = \infty$. $\boldsymbol{\beta}$ is a parameter vector that contains the regression coefficients for the covariate vector $\mathbf{x}_j$. F is a cumulative distribution function which includes the logistic, normal or the extreme value distributions which respectively lead to the logit, probit and complementary log-log link functions. $P(y_j = r)$ is just the difference of the r^{th} and $(r-1)^{th}$ cumulative probabilities.

$$P(y_j = r) = F\left(\alpha_r + \mathbf{x}_j^{'}\boldsymbol{\beta}\right) - F\left(\alpha_{r-1} + \mathbf{x}_j^{'}\boldsymbol{\beta}\right)$$

The link function F^{-1} is a strictly monotonic function in the range [0, 1] onto the real line. The cumulative link model

$$F^{-1}\left[P(y_j \le r)\right] = \alpha_r + \mathbf{x}_j^{'}\boldsymbol{\beta}$$

links the cumulative probabilities to the real line using the link function F^{-1}. This model assumes that effects $\mathbf{x}_j$ are the same for each cutpoint, $r = 1,\dots,R-1$.

For the segmentation model the Proportional Odds model is extended by considering a latent class model with K segments. Let $\boldsymbol{\varphi} = (\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\pi})$ be the vector comprising the parameters of the latent class model with K segments. The n^{th} density function is of the form

$$P(\mathbf{Y}_n = \mathbf{y}_n|\boldsymbol{\varphi}) = \sum_{k=1}^{K} \pi_k.P\left(\mathbf{Y}_n = \mathbf{y}_n|\boldsymbol{\alpha},\boldsymbol{\beta}_k\right)$$

π_k are the proportion of respondents that are allocated to each segment such that $\sum_1^K \pi_k = 1$ and $P(y_{jn} = r|\boldsymbol{\alpha},\boldsymbol{\beta}_k)$ is the Proportional Odds model. The log-likelihood function

$$l(\boldsymbol{\varphi}) = \sum_{n=1}^{N}\ln\sum_{k=1}^{K} \pi_k.P\left(\mathbf{Y}_n = \mathbf{y}_n|\boldsymbol{\alpha},\boldsymbol{\beta}_k\right)$$

is maximized through the EM algorithm. The procedure is similar to the one described for latent class regression models where observed data is augmented by introducing unobserved data λ_{nk} in the complete likelihood function.

$$L(\boldsymbol{\varphi} \mid \boldsymbol{\Lambda}) = \prod_{n=1}^{N}\prod_{k=1}^{K}\left[\pi_k.P\left(\mathbf{Y}_n = \mathbf{y}_n|\boldsymbol{\alpha},\boldsymbol{\beta}_k\right)\right]^{\lambda_{nk}}$$

In the E-step, the expected log-likelihood function is derived by replacing λ_{nk} by posterior probabilities $\hat{p}_{nk}$.

$$E[l(\boldsymbol{\varphi}|\boldsymbol{\Lambda})] = \sum_{n=1}^{N}\sum_{k=1}^{K}[\hat{p}_{nk}.\ln(\pi_k) + \hat{p}_{nk}.\ln P(\mathbf{Y}_n = \mathbf{y}_n|\boldsymbol{\alpha},\boldsymbol{\beta}_k)]$$

In the M-step, the two terms on the right hand side of $E[l(\boldsymbol{\varphi}|\boldsymbol{\Lambda})]$ are maximized separately. The maximization of the first term with respect to π_k, is carried out by the method of Lagrange multipliers subject to the constraint $\sum_1^K \pi_k = 1$. The maximization of the second term with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}_k$ is carried out by transforming the polychotomous responses as a vector of 0-1 indicators, which allows the use of Poisson likelihood in the model fit. Hence each term of $\sum_1^N \sum_1^K p_{nk}.\ln P(\mathbf{Y}_n = \mathbf{y}_n|\boldsymbol{\alpha}, \boldsymbol{\beta}_k)$ is considered as a weighted Poisson log-likelihood function.

SIMULATION STUDY

In order to assess the performance of the two proposed models, synthetic data sets were generated using a GLIM algorithm. The simulation study was devised to mimic the application of Camilleri and Green (2003) in which four car brands; four price values and two door features were generated to define the car attributes. In the application, a full profile design was employed in which 32 items were generated where each item had a unique attribute combination. This guaranteed a full factorial approach.

To simulate subjects' responses, parameter values were required for each of the *K* segments. A set of parameter values was generated for each model described in the preceding section using a data set illustrated in Camilleri and Green (2003). Model 1 is the latent class regression model using mixtures of multivariate conditional normal distributions. Model 2 is the latent class model, which accommodates a Proportional Odds model using a probit link function. The linear predictor used in both models included all main effects and all pairwise interactions of the three car attributes. The parameter vectors $\boldsymbol{\beta}_k$ and $\boldsymbol{\pi}$ were used to simulate rating responses and segment allocation for each of the *N* hypothetical subjects.

To allocate the *N* hypothetical subjects into *K* segments, the known proportions π_k were used to compute the cumulative probabilities, $q_0, q_1, ..., q_K$, where $q_0 = 0$, $q_k = \sum_{i=1}^k \pi_i$ 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 value was 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 32 synthetic data values for each subject, given his segment allocation, the known parameter values $\boldsymbol{\beta}_k$ and the known item attribute values were substituted in the linear predictor. The linear predictors were then perturbed by adding an error terms ε_i to these utility values to have either Logistic, Normal or 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.

$$\varepsilon_i = \log\left(\frac{u_i}{1-u_i}\right) \text{ if } \varepsilon_i \text{ has a Logistic distribution}$$

$$\varepsilon_i = \Phi^{-1}(u_i) \text{ if } \varepsilon_i \text{ have a Normal distribution}$$

$$\varepsilon_i = \log\left[-\log(1-u_i)\right] \text{ if } \varepsilon_i \text{ has an Extreme distribution}$$

A set of 6 specified cutpoint values α_r was used to modify these perturbed linear predictors to rating scores ranging from 1 to 7. Utility values ranging from α_{r-1} to α_r were categorized as having a rating score *r*. This classification gave the rating responses for each hypothetical subject a random category allocation.

To investigate model performance of the two latent class models, a number of data sets were generated by using different sets of values u_i. Four factors that are listed 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
- Size of perturbation
- Distribution of the error terms ε_i

The size of the perturbation was varied by multiplying the error term ε_i by a specified scalar *c*. These four factors, which reflect a variation in conditions in applications, are examined as parameters of the experiment since they are expected to affect model performance.

The following six measures are normally used to assess computational effort, goodness of fit, predictive power, parameter recovery 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.

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

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

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

$\hat{\pi}_k$ 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.

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

$\hat{y}_{nj}$ 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 assessed by each subject. A predicted response for model 2 was set to the rating category with the highest predicted probability.

- 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.

RESULTS OF THE SIMULATION STUDY

A major limitation of the EM algorithm is that it can converge on local stationary points and global maxima are not guaranteed. To overcome this limitation, five data sets were generated for each model varying the type of distribution for ε_i, number of subjects N, number of segments K and the perturbation constant c. These starting values were selected from a wide range of seed numbers. N and K were each varied at two levels (200 and 400 subjects; 2 and 4 segments). Two values were considered for the constant c (0.3 and 1). The Logistic, Normal and Extreme value distributions were considered for ε_i. Each simulated data set was re-fitted using either the latent class model 1 or model 2. $RMS(\hat{\boldsymbol{\beta}})$, $RMS(\hat{\boldsymbol{\pi}})$ and $RMS(\hat{\mathbf{y}})$ 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.

Measure	N	Model 1	Model 2
R^2	200	0.8216	0.8613
	400	0.8056	0.8473
Number of iterations	200	27.4	30.4
	400	28.9	28.8
Segment recovery	200	95.1%	99.0%
	400	97.3%	99.4%
rms $(\hat{\boldsymbol{\beta}})$	200	0.2385	0.2109
	400	0.1679	0.1473
rms $(\hat{\boldsymbol{\pi}})$	200	0.0352	0.0461
	400	0.0309	0.0302
rms $(\hat{\mathbf{y}})$	200	2.9651	2.1538
	400	2.8945	2.0956

Table 1: Model performance by number of subjects

Table 1 displays model performance for the two models by considering 2 segments, a perturbation constant $c = 0.3$, and assuming a Normal distribution for ε_i. An increase in the number of subjects improves parameter recovery; however, goodness of fit deteriorates with an increase in the sample size. A change in the sample size has negligible effect on computational effort and segment membership recovery. A latent class model that accommodates the proportional odds model outperforms the latent class regression model in all the measures except computational effort.

Measure	K	Model 1	Model 2
R^2	2	0.8512	0.8852
	4	0.9033	0.9361
Number of iterations	2	28.3	29.7
	4	35.6	37.4
Segment recovery	2	97.6%	98.3%
	4	94.8%	97.1%
rms $(\hat{\boldsymbol{\beta}})$	2	0.1732	0.1231
	4	0.2536	0.2246
rms $(\hat{\boldsymbol{\pi}})$	2	0.0319	0.0322
	4	0.0168	0.0159
rms $(\hat{\mathbf{y}})$	2	2.1156	1.6897
	4	3.0598	2.6781

Table 2: Model performance by number of segments

Table 2 displays model performance for the two models by considering 400 subjects, a perturbation constant $c = 0.3$, and assuming a Logistic distribution for ε_i. Computational effort increases when fitting a latent class model with more segments. Parameter recovery, predictive accuracy and segment membership recovery deteriorate with an increase in the number of segments; however, goodness of fit and recovery of segment membership probabilities improve by an increase in the number of segments. Model 2 performed better than model 1 in almost all the measures. The number of iterations required and the mean $RMS(\hat{\boldsymbol{\pi}})$ are comparable for both models.

Measure	c	Model 1	Model 2
R^2	0.3	0.8017	0.8223
	1	0.5452	0.5489
Number of iterations	0.3	30.4	30.9
	1	32.2	33.1
Segment recovery	0.3	93.1%	95.3%
	1	73.4%	79.6%
rms $(\hat{\boldsymbol{\beta}})$	0.3	0.2481	0.2134
	1	0.3261	0.3144
rms $(\hat{\boldsymbol{\pi}})$	0.3	0.0316	0.0329
	1	0.1896	0.1529
rms $(\hat{\mathbf{y}})$	0.3	2.8614	2.5319
	1	3.3614	3.2189

Table 3: Model performance by perturbation constant

Table 3 displays model performance for the two models by considering 400 subjects and 2 segments and an extreme value distribution for ε_i. Inevitably, the amount of added error decreases the performance of the algorithms, which is a well-known result in estimation theory. Furthermore, goodness of fit, parameter recovery, segment membership and predictive accuracy recovery deteriorate by increasing the error variance. Once more, model 2 is outperforming model 1 in most of the measures.

Measure	Distribution	Model 1	Model 2
R^2	Logistic	0.9011	0.9216
	Normal	0.9349	0.9356
	Extreme	0.8544	0.8874
Number of iterations	Logistic	35.7	36.4
	Normal	36.9	37.5
	Extreme	35.9	35.8
Segment recovery	Logistic	97.8%	98.2%
	Normal	98.3%	98.2%
	Extreme	90.3%	94.1%
rms $(\hat{\boldsymbol{\beta}})$	Logistic	0.2786	0.2586
	Normal	0.2589	0.2572
	Extreme	0.3325	0.3063
rms $(\hat{\boldsymbol{\pi}})$	Logistic	0.0258	0.0227
	Normal	0.0211	0.0231
	Extreme	0.0312	0.0208
rms $(\hat{\mathbf{y}})$	Logistic	3.0716	3.0511
	Normal	3.0599	3.0435
	Extreme	3.2566	3.1207

Table 4: Model performance by type of distribution

Table 4 displays model performance for the two models by considering 200 subjects, 4 segments and a perturbation constant $c = 0.3$. For each model, the six measures give comparable results when the choice of the distribution is Normal or Logistic. Goodness of fit, parameter recovery, segment membership and predictive accuracy recovery deteriorate when the extreme value distribution is used. The performance of model 1 is at best comparable to model 2 when the distribution of ε_i is Normal. This is due to the fact that model 1 assumes mixtures of multivariate conditional normal distributions.

CONCLUSIONS

A latent class model that accommodates the proportional odds model outperforms the latent class regression model in segment membership and parameter recovery, goodness of fit and predictive accuracy. Goodness of fit improves for fewer data points and more segments; however, parameter recovery, segment membership recovery and predictive accuracy deteriorate with an increase in fitted segments. Computational effort increases with an increase in the number of clusters but is not affected by the choice of distribution, sample size and error variance. The choice of the error distribution has noticeable effect on the performance of the two models. The Normal and Logistic distributions outperform the Extreme value distribution. This may be attributed partly to the similarity of Normal and Logistic distributions and partly to the choice of the model to generate the parameter vectors $\boldsymbol{\beta}_k$ and $\boldsymbol{\pi}$.

REFERENCES

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.

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., Wedel, M., Vriens, M. and Ramaswamy, V. (1992), Latent Class Metric Conjoint Analysis, *Marketing Letters*, 3,3, 273-288.

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

Helsen, K., Jedidi, K. and DeSarbo, W.S. (1992): A new approach to country segmentation using multinational diffusion patterns. *Journal of Marketing*, 57: 60-71

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

Ramaswamy, V., DeSarbo, W.S., Reibstein, D., Robinson (1993), An Empirical Pooling Approach to Estimate Marketing Mix Elasticities with PIMS Data, *Marketing Science*, 12, 103-124.

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.

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.

View publication stats