

## An Application of Latent Class Random Coefficient Regression

LARS ERICHSEN<sup>†</sup> lsen@novonordisk.com  
*Pharmacokinetics, Novo Nordisk, Novo Nordisk Park G8.2.30, 2670 Maaloev, Denmark*

PER BRUUN BROCKHOFF pbb@imm.dtu.dk  
*Informatics and Mathematical Modelling, Richard Petersens Plads, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark*

**Abstract.** In this paper we apply a statistical model combining a random coefficient regression model and a latent class regression model. The EM-algorithm is used for maximum likelihood estimation of the unknown parameters in the model and it is pointed out how this leads to a straightforward handling of a number of different variance/covariance restrictions. Finally, the model is used to analyze how consumers' preferences for eight coffee samples relate to sensory characteristics of the coffees. Within this application the analysis corresponds to a model-based version of the so-called external preference mapping.

**Keywords:** Latent class regression, random coefficient regression, principal component regression, finite mixture distributions, bootstrapping, maximum likelihood estimation, EM-algorithm, sensory analysis, consumer preferences, preference mapping.

### 1. Introduction

Most statistical models are based on some homogeneity assumptions about the parameters of interest, one important example being the traditional regression model, where regression coefficients are assumed constant across observations.

If observations are on several subjects, a random coefficient regression (RCR) model, Longford [13], is frequently used to account for correlation between observations on the same subject, the target being an average regression relation over the population rather than individual relations. The assumption is then that subjects are drawn randomly from the same homogeneous population.

In the present paper the emphasis is on estimating regression parameters of a heterogeneous population. This may be done within the framework of normal linear models if a partitioning of the subjects, into a priori known

---

<sup>†</sup> Requests for reprints should be sent to Lars Erichsen, Pharmacokinetics, Novo Nordisk, Novo Nordisk Park G8.2.30, 2670 Maaloev, Denmark.

homogeneous groups, is available, and the analysis is carried out conditionally on the variable defining the groups. If no such variable is known, we may take it into our model as a latent (unobserved) random variable, i.e. assume the existence of a finite (fixed) number of latent classes/segments, yielding a family of mixture distributions, which is denoted a latent class regression (LCR) model, Wedel and Desarbo [23], or a switching regression model, Kiefer [11], in the literature.

We use a latent class random coefficient regression model, which is a combination of a random coefficient regression and a latent class regression model, to be used in a regression setup with repeated measurements on subjects from a heterogeneous population. Maximum likelihood estimates are obtained iteratively using the expectation-maximization (EM)-algorithm of Dempster et al. [5]. The latent class mixed effects model is presented by Verbeke and Lesaffre [22] and they also mention the EM algorithm, so the present work can be seen as a model specific implementation and application of the more general setting of Verbeke and Lesaffre [22].

We apply the model to empirical data from a tasting experiment consisting of a sensory study and a consumer survey with coffee. In the sensory study different coffee products were evaluated by trained assessors with respect to a number of sensory attributes, each assessor assigning scores to each product. The consumer survey was conducted with the same coffee products and each consumer was asked to give a preference score to each product. The aim is to explain the consumer preferences by the sensory attributes in a regression type setup with preference as the response and sensory variables as predictors, dividing consumers into segments according to which sensory profiles they prefer.

## 2. Specification of the Model

The LCRCR model outlined in the introduction is specified as a mixture of mixed linear models, the mixed part stemming from random subject effects. Further restrictions on the variance/covariance parameters are discussed.

### 2.1. The latent class random coefficient regression model

Let  $y_{ij}$  denote the  $j$ 'th observation of subject  $i$  on a single scalar response and  $x_{ij}$  the corresponding  $p \times 1$  design/covariate matrix for  $i = 1, \dots, I$  and  $j = 1, \dots, n_i$ . Further  $z_i$  denotes the random variable taking the value  $s$  if subject  $i$  belongs to segment  $s$ ,  $s = 1, \dots, S$ .

The LCRCR-model is now specified by the statements (1)-(5) below:

$$y_{ij} = x'_{ij}B_i + \epsilon_{ij} \tag{1}$$

$$B_i|z_i = s \sim N_p(\beta_s, \Sigma_s) \tag{2}$$

$$\epsilon_{ij}|z_i = s \sim N(0, \sigma_s^2) \tag{3}$$

together with the assumption of mutual independence of  $z_1, \dots, z_I$  and conditional mutual independence of all the variables  $B_i$  and  $\epsilon_{ij}$  given the  $z_i$ 's. We have used the notation  $B_i|z_i = s$  to indicate the conditional distribution of  $B_i$  given that  $z_i = s$  and  $N_p(\beta, \Sigma)$  to denote the  $p$ -dimensional normal distribution with mean  $\beta$  and covariance matrix  $\Sigma$ .

The  $z_i$ 's are assumed to arise from the same distribution with proportions

$$P(z_i = s) = \pi_s, \quad s = 1, \dots, S \tag{4}$$

The unknown parameters of the model are  $\theta_s = (\pi_s, \psi_s)$ ,  $s = 1, \dots, S$ , where  $\psi_s = (\beta_s, \Sigma_s, \sigma_s^2)$ , and the parameter space is given by

$$\beta_s \in \mathbf{R}^p, \quad \sigma_s^2 > 0, \quad \Sigma_s \text{ is positive definite}, \quad 0 < \pi_s < 1, \quad \sum_{s=1}^S \pi_s = 1 \tag{5}$$

Note that the LCR-model corresponds to the case of  $\Sigma_s = 0$ .

If we collect all responses of subject  $i$  in  $y_i = (y_{i1}, \dots, y_{in_i})'$  and let  $X_i$  denote the  $n_i \times p$ -matrix with rows  $x'_{i1}, \dots, x'_{in_i}$ , the conditional density,  $f_{i|s}(\cdot, \psi_s)$ , of  $y_i$  given  $z_i = s$ , is an  $n_i$ -dimensional multivariate normal density with mean  $X_i\beta_s$  and covariance matrix  $\Sigma_{is}$ , where

$$\Sigma_{is} = X_i\Sigma_sX'_i + \sigma_s^2I_{n_i} \tag{6}$$

and  $I_{n_i}$  is the  $n_i$ -dimensional identity matrix.

The unconditional mean and covariance matrix of  $y_i$  are given by

$$E(y_i) = \sum_{s=1}^S \pi_s X_i\beta_s \tag{7}$$

and

$$Var(y_i) = \sum_{s=1}^S \pi_s [\{X_i\Sigma_sX'_i + \sigma_s^2I_{n_i}\} + AA'] \tag{8}$$

where

$$A = X_i\beta_s - \sum_{s=1}^S \pi_s X_i\beta_s,$$

but the distribution of  $y_i$  is not normal if the subjects divide into at least two different segments.

## 2.2. Covariance structure

One may wish to restrict the variance/covariance parameters in order to make estimation feasible in cases with sparse data or to avoid segmentation on the basis of variance parameters when the mean parameters are the target.

The restrictions are divided into two essentially different sets of restrictions, the first concerning the distribution of the random coefficients within segments, the second concerning assumptions of variance/covariance homogeneity across segments.

For the random coefficient distribution the following three cases are considered:

1.  $\Sigma_s$  positive definite (unrestricted).
2.  $\Sigma_s = \text{diag}(\sigma_{s1}^2, \dots, \sigma_{sp}^2)$ .
3.  $\Sigma_s = \sigma_s^2 I_p$ .

The different variance/covariance homogeneity assumptions yield four levels of restrictions:

- A No restrictions across segments.
- B  $\sigma_1^2 = \dots = \sigma_S^2 = \sigma^2$ .
- C  $\Sigma_1 = \dots = \Sigma_S = \Sigma$ .
- D  $\sigma_1^2 = \dots = \sigma_S^2 = \sigma^2$  and  $\Sigma_1 = \dots = \Sigma_S = \Sigma$ .

## 2.3. Nuisance parameters

So far, we have included all mean parameters in the segmentation of the subjects. However, there may be parameters that are considered nuisance parameters and therefore are not a target for the segmentation. One example of nuisance parameters would be parameters describing subject specific levels in cases where we are interested in how changes of covariates affect the response.

We consider the modification of (1)

$$y_i = D_i \xi_i + X_i B_i + \epsilon_i \quad (9)$$

where  $D_i$  is a known injective design matrix and  $\xi_i \in \mathbf{R}^{q_i}$  is a nuisance parameter.

### 3. Estimation of the Parameters

In the following sections we discuss identifiability and maximum likelihood estimation of the parameters.

#### 3.1. Identifiability

The LCRCR model is only identifiable up to permutation of the segment labels  $1, \dots, S$ , but this problem may be overcome by imposing the restrictions

$$\pi_1 \leq \pi_2 \leq \dots \leq \pi_S \quad (10)$$

on the mixing proportions (see Aitken and Rubin [1]). In practice the estimation may be carried out without this restriction, doing the reordering post hoc. Of course any ordering of the segment labels is as good as any other and we follow McLachlan and Basford [16], reporting the parameter estimates of one of the  $S!$  permutations.

Yakowitz and Spragins [25] show that a mixture of unrestricted multivariate normal distributions is identifiable. Identifiability in the LCRCR case is thus ensured if the parameters of the normal components are identified. In the balanced case with all  $X_i = X$ , the means and covariance matrices are  $X\beta_s$  and  $X\Sigma_s X' + \sigma_s^2 I$  respectively, and the model is identifiable if and only if  $X$  is injective and  $n = n_i > p$ . If  $n = p$  the mean parameters of the normal components are still identifiable, but this is not the case for the variance parameters  $\sigma_s^2$  unless further restrictions are imposed on  $\Sigma_s$ . In the unbalanced case the identifiability problem is more complicated but identifiability is ensured if the above conditions hold true for at least one subject, though this is no longer a necessary requirement in general.

Titterton et al. [21] give an introduction to identifiability of mixtures and references to further literature on the subject.

#### 3.2. The likelihood function

By the independence assumptions we obtain the following likelihood function for all observations:

$$L_y(\theta) = \prod_{i=1}^I \sum_{s=1}^S \pi_s f_{i|s}(y_i, \psi_s) \quad (11)$$

Maximum likelihood estimation may now be carried out by direct maximization of (11), e.g. using the Newton-Raphson algorithm or Fisher's scoring method. This requires the calculation of the first and second order derivatives of the log likelihood function and we will not go through with this, but use the more easily implemented EM-algorithm instead.

It is well known that mixture distributions may have unbounded likelihood functions, see e.g. Day [4]. For a subject with  $y_i = X_i\beta_s$  the likelihood function increases unboundedly as the determinant of the covariance matrix in segment  $s$  tends to 0. Thus maximum likelihood estimation in the usual sense breaks down. In many cases, however, a local maximum point of the likelihood function provides a consistent estimate of the parameters. Hathaway [10] discusses regularity conditions under which a local maximum likelihood estimator exists and is consistent, while Kiefer [11] considers the special case of the switching regression model.

### 3.3. The EM-algorithm

The EM-algorithm is an iterative optimization algorithm and was formally developed by Dempster et al. [5] with the purpose of optimizing likelihood functions in incomplete data situations. Incomplete data was to be understood in a broad sense including missing observations and unobservable random variables. Hence, the EM-algorithm is commonly used for mixture models, see e.g. Redner and Walker [19] and Desarbo and Cron [6]. The details of the algorithm for the specific model of the present paper are therefore omitted.

The fitting of the models using the various variance/covariance restrictions of section 2.2 can be seen to be simple modifications of the M-step of the algorithm.

Regarding the handling of nuisance parameters, instead of estimating  $\xi_i$  in each step of the algorithm, transformation of the data beforehand can be done to eliminate the nuisance parameters: if  $Q_i$  is an  $n_i \times (n_i - q_i)$ -matrix with orthonormal columns and  $Q_i'D_i = 0$ , the distribution of the transformed data,  $Q_i'y_i$ , does not depend on the nuisance parameters and the original EM-algorithm may be used with  $y_i$  replaced by  $Q_i'y_i$  and  $X_i$  by  $Q_i'X_i$ . This saves computation time, since the modification is done only once.

#### 4. Application to Coffee Data

The present data is part of a larger data set from a tasting experiment with coffee, described in more detail in ESN [9]. This part of the data set was used for a workshop at the Second Pangborn sensory science symposium, Davis, USA, 1995 and workshop contributions are published, see e.g. Brockhoff [3].

In a sensory study 45 trained assessors evaluated 8 different coffee products with respect to 16 sensory properties on the scale 0, 10, ..., 100, a high score indicating a high level of the given sensory attribute. Each assessor tasted the same product three times, giving a total of  $3 \times 45 \times 8 = 1080$  observations for each sensory property. The sensory properties were acid taste, bitter taste, burnt flavor, caramel flavor, wood flavor, chemical flavor, earth flavor, astringent mouth feel, floral flavor, fruity flavor, grass flavor, malty flavor, rancid flavor, salty taste, sweet taste, and sour taste.

In addition to the sensory study, data from a consumer survey with the same eight products were available. In this survey consumers were asked to assign a preference score on the scale 1, 2, ..., 9 to each product, a high score indicating a high preference of the given product.

The sensory evaluations were conducted by the International coffee organization in England and the consumer survey involved 369 consumers from the five countries: Denmark (51), France (80), Germany (73), Poland (80), and the UK (85).

The sensory variables were standardized to have error variance one, where the error variance was estimated from the residuals in a two-way analysis of variance, using assessor and product as factors (including interaction) and the seven principal components of the standardized sensory variables were used as predictors in a regression setup with consumer preferences as responses. The standardization of the sensory variables was done in order to avoid principal directions based on random error variation.

We fitted the LCRCR-model with restrictions 2 D (see section 2.2) to the data, after eliminating the consumer levels according to section 2.3. The results of testing the effects of the PC-scores in the presence of 1, 2 and 3 segments, respectively, are seen in Table 1 and tests of the number of segments being 1 or 2 against the alternatives 2 and 3 segments, respectively, are given in Table 2. The latter two tests were carried out by bootstrapping (see McLachlan [15]) since a hypothesis of having one segment less corresponds to a segment proportion equal to 0, which is a boundary value of the parameter space, and the usual asymptotic theory for the likelihood ratio statistic does not apply (see McLachlan and Basford [16], pp. 21). In each case 100 bootstrap samples were generated by simulation, using

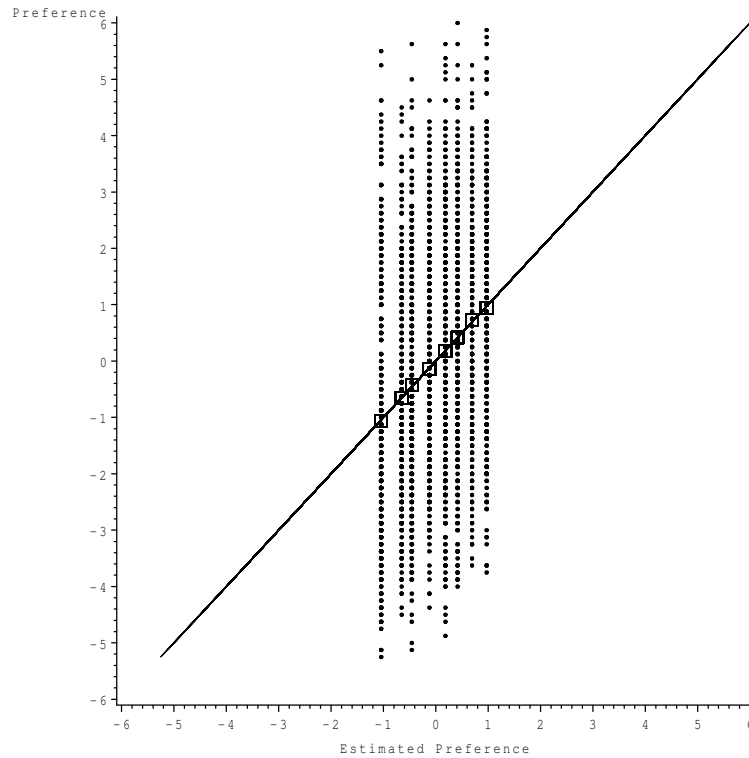


Figure 1. Centered preferences and estimates - no segmentation.

the parameter estimates under the hypothesis to be tested. The likelihood ratio statistic was calculated for each of the 100 bootstrap samples and the fifth largest used as an estimate,  $\hat{q}_{0.95}$ , of the 95 % quantile in the distribution of the statistic. Several different starting values for the parameters were used in each case to be as sure as possible that a maximum of the likelihood function was obtained.

Summarizing the test results, only the fourth principal component is insignificant in the cases with 1, 2 or 3 segments. A test of the significance of the fourth principal component was done against the full model whereas the other tests reported in Table 1 are tests against the alternative model with the fourth component removed. From the test results in Table 2, we conclude that the 1-segment model describes the data inadequately whereas the 2-segment model is accepted when tested against the 3-segment model.



Table 1. Likelihood ratio tests for exclusion of each principal component within a model with one, two and three segments respectively.

cov	1 segment			2 segments			3 segments		
	-2logQ	df	P (%)	-2logQ	df	P (%)	-2logQ	df	P (%)
pc1	82.5	1	0.0	145.8	2	0.0	160.5	3	0.0
pc2	17.8	1	0.0	88.9	2	0.0	96.8	3	0.0
pc3	28.6	1	0.0	29.7	2	0.0	34.4	3	0.0
pc4	0.4	1	52.7	2.6	2	27.3	4.8	3	18.7
pc5	176.5	1	0.0	239.8	2	0.0	254.5	3	0.0
pc6	41.0	1	0.0	48.0	2	0.0	55.7	3	0.0
pc7	12.3	1	0.0	13.2	2	0.1	16.8	3	0.1

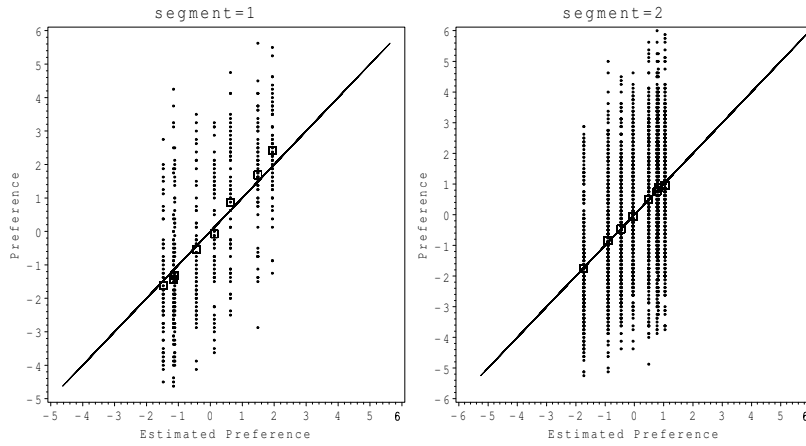


Figure 2. Centered preferences and estimates - 2 segments.

Table 2. Tests of the number of segments by bootstrapping. The second column is the 95% quantile of the statistic in 100 bootstrap samples.

#segments	-2logQ	$\hat{q}_{0.95}$
1	63.2	21.7
2	14.7	19.7

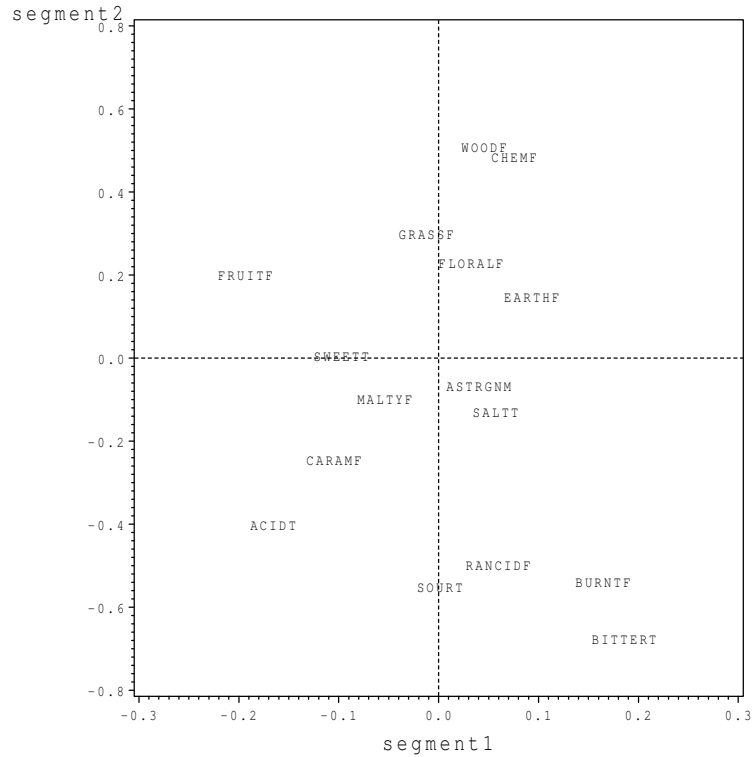


Figure 3. Regression coefficients of sensory attributes.

Plots of the centered preferences against the estimated means for the 1- and 2-segment models are shown in Figures 1 and 2. From these it is evident that the signal to noise ratio is low. However the 2-segment model fits the data noticeably better than the 1-segment model. The estimates of the segment proportions yield a second segment with a little more than 80% of the consumers.

The characteristics of the two segments in terms of the original (scaled) sensory variables are illustrated in Figure 3. We see that bitter taste and burnt flavor seem to be well liked in segment 1 but disliked in segment 2 whereas acid taste seems to be disliked by consumers from both segments.

If a consumer is assigned to the segment with the highest posterior, consumer specific segmentation probability, we can look at the empirical distribution of these probability estimates to get an idea of how well the

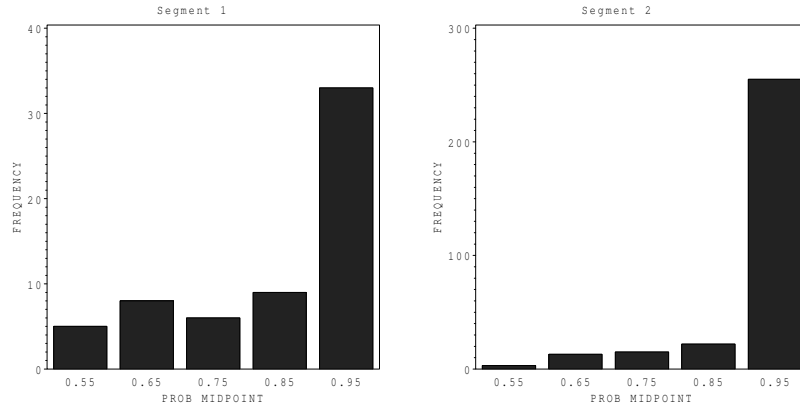


Figure 4. Empirical distribution of segmentation probabilities.

segmentation has worked. The histogram in Figure 4 shows the number of consumers with segmentation probabilities in intervals of length 0.1. From this it seems that the segmentation into segment 2 is more certain than the segmentation into segment 1 in general. A larger proportion of the consumers in segment 1 seem to be genuine mixtures of the two segments than is the case in segment 2. Overall around 80% of the consumers were assigned to a segment with probability greater than 0.9.

Finally, Figure 5 shows the distribution of the consumers on the segments for each country and we see that the three countries Denmark, France and Germany appear to be more homogeneous with respect to the coffee preferences than the populations in Poland and the UK. In all five countries, however, the majority of the consumers are assigned to the second segment. The population in segment 1 is largely constituted by a group of Poles and UK consumers, who apparently have a different preference for coffee than the rest of the consumers.

## 5. Discussion

The EM algorithm has previously been used for maximum likelihood estimation in linear mixed models, Laird et al. [12], and in latent class regression, Desarbo and Cron [6], and we have shown that the algorithm may be applied to a model, which combines the latent class regression and the

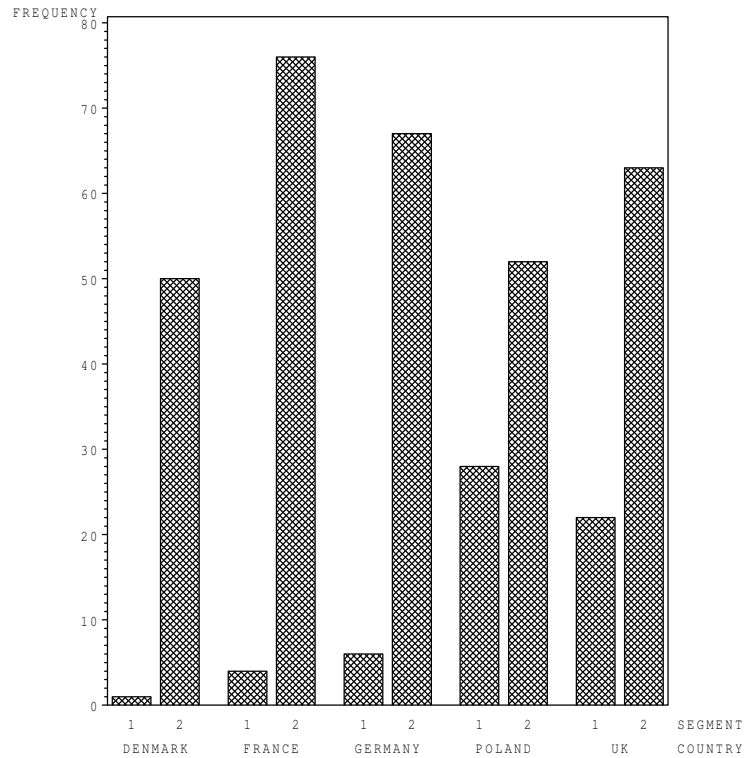


Figure 5. Distribution of subjects on segments for each country.

mixed model, without losing the tractable property of having closed form solutions to the E- and M-steps.

The latent class random coefficient regression model does not take measurement error in the covariates into account. Furthermore a discrete version of the model would be appropriate in the analysis of the coffee data, possibly based on threshold type models. Muthén and Shedden [17] present a random effects mixture model for multiple binary outcome, while Wedel and Desarbo [24] consider mixtures of generalised linear models.

It is straightforward to extend the distribution of the latent class variables to incorporate information about the subjects. In the case of the consumer data, we could allow segment proportions to depend on country, thus using the information of subjects' nationality. The segmentation for each nationality was deduced post hoc in our data analysis (section 4).

Arminger et al. [2] consider mixtures of multivariate normal distributions. In cases where the covariance matrix is restricted, they suggest estimating the parameters of the corresponding unrestricted model and then estimate the parameters in the restricted case by minimizing some measure of distance between the two sets of covariance parameters.

Besides the likelihood approach, Bayesian analyses of mixture densities are widely used in the literature, see e.g. Diebolt and Robert [7], Escobar and West [8] and Richardson and Green [20]. Markov chains are used, together with a priori distributions for the number of mixture components as well as for the mixture component parameters, to deal with the problem of having an unknown number of segments, enabling inference about this number. However, the choice of prior distributions is somewhat controversial and has an impact on the final inference about the number of segments.

An application of a restricted version (D3) of the LCRCR model is given in Poulsen et al. [18] to model the relation between consumer preferences and sensory characteristics of frozen peas. In Poulsen et al. [18] it is also outlined how the model can be seen as a model based so-called 'External preference mapping', see McEwan [14].

## References

1. M. Aitken and D. B. Rubin. Estimation and hypothesis testing in finite mixture models. *Journal of the Royal Statistical Society, Series B*, 47: 67-75, 1985.
2. G. Arminger, P. Stein and J. Wittenberg. Mixtures of conditional mean and covariance structure models. Unpublished working paper, 1995.
3. P. M. Brockhoff. Growth curve modeling. *Food Quality and Preference*, 9(3): 91-93, 1998.
4. N. E. Day. Estimating the components of a mixture of normal distributions. *Biometrika*, 56:463-474, 1969.
5. A. P. Dempster, N. M. Laird and R. B. Rubin. Maximum likelihood from incomplete data via the EM-algorithm. *Journal of the Royal Statistical Society, Series B*, 39:1-38, 1977.
6. W. S. Desarbo and W. L. Cron. A maximum likelihood methodology for clusterwise regression. *Journal of classification*, 5:249-282, 1988.
7. J. Diebolt and C. Robert. Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society, Series B*, 56:163-175, 1994.
8. M. D. Escobar and M. West. Bayesian density estimation and inference using mixture distributions. *Journal of the American Statistical Association*, 90:577-588, 1995.
9. ESN. A European sensory and consumer study. A case study on coffee. Collaboration by the European Sensory Network, 1996.
10. R. J. Hathaway. A constrained formulation of maximum likelihood estimation for normal mixture distributions. *Annals of Statistics*, 13:795-800, 1985.

11. N. M. Kiefer. Discrete parameter variation: Efficient estimation of a switching regression model. *Econometrica*, 46(2): 427-434, 1978.
12. N. Laird, N. Lange and D. Stram. Maximum likelihood computations with repeated measures: Application of the EM algorithm. *Journal of the American Statistical Association*, 82:97-105, 1987.
13. N. T. Longford. *Random coefficient models*. Clarendon press, Oxford, 1993.
14. J. A. McEwan. Preference Mapping for product optimization. In *Multivariate Analysis of Data in Sensory Science*, eds. T. Næs and E. Risvik. Elsevier Science, Amsterdam, 1996.
15. G. J. McLachlan. On bootstrapping the likelihood ratio statistic for the number of components in a normal mixture. *Applied Statistics*, 36(3): 318-324, 1987.
16. G. J. McLachlan and K. E. Basford. *Mixture Models: Inference and Applications to Clustering*. New York, N. Y: Marcel Dekker, Inc, 1988.
17. B. Muthén and K. Shedden. Finite mixture modelling with mixture outcomes using the EM algorithm. *Unpublished paper*, 1998.
18. C. S. Poulsen, P. M. Brockhoff and Erichsen, L. Heterogeneity in consumer preference data - a combined approach. *Food Quality and Preference*, 8(5/6): 409-417, 1997.
19. R. A. Redner and H. F. Walker. Mixture densities, maximum likelihood and the EM-algorithm. *SIAM Review*, 26(2) 195-239, 1984.
20. S. Richardson and P. J. Green. On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society, Series B*, 59(4):731-792, 1997.
21. D. M. Titterton, A. F. M. Smith and U. E. Makov. *Statistical analysis of finite mixture distributions*. New York, N. Y: John Wiley and Sons, 1985.
22. G. Verbeke and E. Lesaffre. A linear mixed-effects model with heterogeneity in the random effects population. *Journal of the American Statistical Association*, 91:217-221, 1996.
23. M. Wedel and W. S. Desarbo. A review of recent developments in latent class regression models. In: *R. P. Bogozzi (Ed. ). Advanced methods for marketing research*, 353-388, 1994.
24. M. Wedel and W. S. Desarbo. A mixture likelihood approach for generalized linear models. *Journal of Classification*, 12:21-55, 1995.
25. S. J. Yakowitz and J. D. Spragins. On the identifiability of finite mixtures. *Annals of Mathematical Statistics*, 39:209-214, 1968.