**Random Component Threshold Models in a Customer Satisfaction Evaluation**

**Ayoub Saei, Soroush Alimoradi**

## Abstract

The degree of customer satisfaction is measured on an ordinal scale in evaluating a customer opinion programme. Two random component threshold models are fitted to the results data. Estimation of the parameters in the models and variance components are given by residual maximum likelihood method. The predicted values of the probability of selecting a specific response category are given for all customers. A threshold is selected and customers are then divided into happy and unhappy groups.

**S$^3$RI Methodology Working Paper M04/06**

# Random Component Threshold Models in a Customer Satisfaction Evaluation

**Ayoub Saei**

Southampton Statistical Sciences Research Institute, University of Southampton, Highfield

Southampton SO17 1BJ, UK

**Soroush Alimoradi**

Department of Mathematical Sciences, Isfahan University of Technology, Isfahan 84154, Iran

**Summary:** The degree of customer satisfaction is measured on an ordinal scale in evaluating a customer opinion programme. Two random component threshold models are fitted to the results data. Estimation of the parameters in the models and variance components are given by residual maximum likelihood method. The predicted values of the probability of selecting a specific response category are given for all customers. A threshold is selected and customers are then divided into happy and unhappy groups.

*Keywords*: Ordinal; Question dependent; Question independent; REML; Threshold

## 1. Introduction

The response variable in evaluating a customer opinion programme usually measures the degree of customer's satisfaction based on some categories. The categories are usually ordinal scales, e.g., excellent, good, not bad, and bad. The response category may depend on the data selection method (face-face or mail), customer sub-activities and production lines in a firm or company. The programme is often aimed at finding out what factors influence customers to make particular response category choice. The programme is also aimed at assessing the firm and identifying customers who are or are not satisfied by product and services.

The customer's opinion on product and services is usually a multiple response (responses to the different questions). Thus, model needs to account for variations between and within customers. The aim of this paper is to relate the response variable to the random customer effect acting in the presence of some other covariates and factors. This paper shows how random component threshold models can provide a very good approach to the analysis of the customer opinion data.

Random component threshold models are fitted by Gianola and Foully (1983) and Harville and Mee (1984) using a Bayesian approach. A numerical inte-gration approach along with the EM algorithm is proposed by Jansen (1990 and 1992). The generalized linear mixed models are proposed by Schall (1991), Breslow and Clayton (1993) Wolfinger (1993), McGilchrist (1994), Nelder and Lee (1996) and Lee and Nelder (2001a, 2001b). The model fitting approaches are applied to the random component threshold models by Zhaorong et al (1992) and Saei and McGilchrist (1996), Crouchley (1995), Ten Have (1996) for multivariate ordered response data and Saei and McGilchrist

(1998) to longitudinal ordinal response data.

Section 2 explains the customer satisfaction programme. The models and the estimation approach are given in section 3. Section 4 proposes a predicttion method for selecting a particular response category. Results of application of the methods to the customer opinion data are presented in section 5. The last section provides discussion on the method and results.

## 2. Customer Satisfaction Study

Mobarekeh Steel Complex (MSC) is the main supplier of steel products in Iran and one of the major suppliers in the region. The market research of the MSC sent out questionnaires to evaluate the customer's opinion on product and services. A total of 202 national customers within 9 sub-activities were randomly assigned to two types of data collection (face to face and mailing). The degree of satisfaction of each customer was recorded according to an ordinal response scale (1 = very good, 2 = good, 3 = not bad, 4 = bad and 5 = very bad). There are three different production lines at MSC, hot rolling, cold rolling and pickling (washing Acid). The customers from 9 different sub-activities are used products from one, two or all three production lines. The sub-activities are pipe and profile makers (1), tank, cylinder and container makers (2), home accessories makers (3), steel industries (4), other user of steel sheet (5), heavy steel equipment (6), steel sheet cooperation users (7), motor vehicle makers (8), water, oil and gas pipe makers (9). The data set is available from first author.

## 3. Models and Estimation

The response variable Y is an ordinal random variable, which is assumed to arise from grouping an underlying continues random variable. Let $Y_{is}$ to be the value of ordinal response variable Y on the $s$th question ($s$ = 1, 2,…, 13) for customer $i$, $i$ = 1, 2,..., 202; $Y_{is}$ can take on values 1, 2, ..., 5. The distribution of $Y_{is}$ depends on the linear predictor $\eta_{is}$ in which involves a vector of $p$ known regression variables $\mathbf{x}_{is}$ and the customer random effect $u_i$. Two random component threshold models are defined. The threshold parameters (cut-points) are assumed constant over the questions in the first model. This is called a *question independent random component threshold model*. The model is

(3.1) $P(Y_{is} \leq k = G(\theta_k - \eta_{is})$

where $\theta_k$ are threshold parameters for $k$ = 1, 2,..., 5. The G(.) is a cumulative distribution function for the unobservable continuous random variable $V_{is}$ with conditional mean $\eta_{is}$ and examples of G are given in McCullagh and Nelder (1989). Here the model is a proportional odds model in which G is the cumulative distribution function for a logistic distribution. Results for other models such the proportional hazards model or standard threshold model considered by McCullagh and Nelder (1989) are consistent with the proportional odds model.

The threshold parameters are dependent on the questions in the second model. This model replaces $\theta_k$ by $\theta_{ks}$ in (3.1) and it is called a *question dependent random component threshold model.* The model is

(3.2) $P(Y_{is} \leq k = G(\theta_{ks} - \eta_{is})$

where $\theta_{ks}$ is the threshold parameter for question $s$. Because of the small number of observations in category 5, the categories 5 (very bad) and 4 (bad) are combined, reducing the number of categories under consideration to 4. Thus observation for customer $i$ on question $s$ ( $i$ = 1, 2, …, 202; $s$ = 1, 2,…, 13) is coded as 1 = very good, 2 = good, 3 = not bad and 4 = bad; $k$ = 1, 2, 3, 4. For the model (3.1), the parameter $\theta_0$ is always taken as $-\infty$ so that $G(\theta_0 - \eta_{is})$ is always zero while $\theta_4$ is taken to be $+\infty$ indicating that $G(\theta_4 - \eta_{is})$ is always 1. The parameters $\theta_{s0}$

and $\theta_{s4}$ are reacted as $\theta_0$ and $\theta_4$ in the model (3.2). The threshold parameters are collected into a vector $\boldsymbol{\theta}$.

We consider two different models of the linear predictor $\eta_{is}$ in (3.1) and (3.2). The first model assumes $\eta_{is}$ to be a linear function of a vector $\mathbf{x}_{is}$ of $p$ covariates as well as a random customer effects $u_{1i}$ to account for variation that is not explained by the values in $\mathbf{x}_{is}$. That is

(3.3) $\eta_{is} = \mathbf{x}'_{is}\boldsymbol{\beta} + u_{1i}$

where $\boldsymbol{\beta}$ is an unknown vector of the regression coefficients. The customer random effects $u_{1i}$ are independent normal variables with zero means and variances $\varphi_1$ for $i = 1, 2,\ldots, 202$.

In the second model $\eta_{is}$ includes an extra random effect $u_{2i}$, allowing a possible change in variance and pattern of association on the second and following questions for a customer. This is consistent with the idea that the customers respond differently to the questions. This model is

(3.4) $\eta_{is} = \mathbf{x}'_{is}\boldsymbol{\beta} + u_{1i} + \delta_s u_{2i}$

where $\delta_s = 0$, $s = 1$; $\delta_s = 1$, $s = 2, 3, \ldots, 13$ and $u_{1i}$ and $u_{2i}$ are normal variables with zero means and variances $\varphi_1$ and $\varphi_2$ respectively and covariance between them $\varphi_3$. The $u_{1i}$ is the customer effect at baseline (on the first question), and $u_{2i}$ measures the average deviation from that value on the second and following questions. It is important to realise that a negative value for $\hat{u}_{2i}$ implies a lager decline in $\eta_{is}$, so that lower (very good, good) Y observations are likely.

Let $\mathbf{u}_1 = (u_{11}, u_{12}, \ldots, u_{1N})'$ and $\mathbf{u}_2 = (u_{21}, u_{22}, \ldots, u_{2N})'$ be the customer random effect vectors at baseline (on the first question) and on the second and following questions. In general the model for $\boldsymbol{\eta} = (\eta_{is}, i = 1, 2,\ldots,N = 202, s = , 2, \ldots, n_i = 13)$ can be expressed as $\eta = \mathbf{X}\beta + \mathbf{Z}\mathbf{u}$, where $\mathbf{X}$ is a known matrix of the regression variables, $\mathbf{Z} = \mathbf{Z}_1$ and $\mathbf{u} = \mathbf{u}_1$ under (3.3) and $\mathbf{Z} = [\mathbf{Z}_1,$

$\mathbf{Z}_2$] and $\mathbf{u} = (\mathbf{u}_1', \mathbf{u}_2')'$ under (3.4). The random vectors $\mathbf{u} = \mathbf{u}_1$ and $\mathbf{u} = (\mathbf{u}_1', \mathbf{u}_2')'$ are distributed as multivariate normal with zero mean vectors and variance-covariance matrices $\mathbf{A} = \varphi_1 \mathrm{I}_N$ and $\mathbf{A} = \begin{bmatrix} \varphi_1 \mathbf{I}_N & \varphi_3 \mathbf{I}_N \\ \varphi_3 \mathbf{I}_N & \varphi_2 \mathbf{I}_N \end{bmatrix}$ respectively, where $\mathbf{I}_N$ is an identity matrix of order $N = 202$. An outline of the estimation method is given in Appendix A.

## 4. Customer Response Prediction

Individual reaction on the products and services by MSC is an important factor in the customer satisfaction study. The notion of sensitive customers is also important to the MSC. The prediction approach here is an extension of Anderson and Albert (1981) to include random effects in the linear predictor.

The observable variable $\mathrm{Y}_{is}$ is categorised value of unobservable continuous variable $\mathrm{V}_{is}$ with conditional mean of $\eta_{is}$, i.e., $\mathrm{Y}_{is} = k \Leftrightarrow \theta_{k-1} - \eta_{is} < \mathrm{V}_{is} - \eta_{is} \leq \theta_k - \eta_{is}$. This yields $\mathrm{P}(\mathrm{Y}_{is} = k) = \mathrm{G}(\theta_{k-1} - \eta_{is}) - \mathrm{G}(\theta_{k-1} - \eta_{is})$ and $\mathrm{P}(\mathrm{Y}_{is} = k) = \mathrm{G}(\theta_{sk} - \eta_{is}) - \mathrm{G}(\theta_{s(k-1)} - \eta_{is})$ under models (3.1) and (3.2) respectively. The G(.) is a cumulative distribution for the unobservable continuous variable $\mathrm{V}_{is}$ and given in section 3. The predicted $\hat{\eta}_{is}$ and estimated values $\hat{\theta}_{sk}$ allow a predicted value for $\mathrm{Y}_{is}$, say $\hat{\mathrm{Y}}_{is}$,

$$\hat{\mathrm{Y}}_{is} = k \Leftrightarrow \hat{\theta}_{s(k-1)} < \hat{\eta}_{is} \leq \hat{\theta}_{sk}$$

under the model (3.2). The cumulative probability of second category response, $\mathrm{P}(\hat{\mathrm{Y}}_{is} \leq 2) = \mathrm{G}(\hat{\theta}_{sk} - \hat{\eta}_{is})$, is then selected as a threshold in identifying happy and unhappy customers of the MSC.

## 5. Results

Table 1 shows the parameter estimates and associated standard errors for four question dependent threshold models fitted to the customer satisfaction data. Model 1 is a fixed effect threshold model, fitted via ML. Model 2 is a one random component threshold model, fitted via REML. Models 3 and

4 are threshold models with independent and dependent random components respectively. These models are also fitted via REML. The significance of the parameters in the models can be judged by comparing the estimated values with their asymptotic standard errors. It is important to realize that a negative coefficient for a regression variable in $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$ means that $\boldsymbol{\eta}$ is decreased for an increase in **X** component values, so that $\boldsymbol{\theta} - \boldsymbol{\eta}$ values are increased and lower **Y** observations are likely. This is also true for predicted values of **u**, i.e., a negative predicted value of the random effect further shifts $\boldsymbol{\eta}$ to the left and increases the probability of observing a lower value (very good, good) for **Y**.

The results show that there is a statistically significant variation between customers with REML estimated value $\hat{\varphi}_1$= 0.94 and standard error of 0.44. The results also show that the customer effect does significantly change over questions. The REML estimate of the variance of carrying over question random effect is $\hat{\varphi}_2$= 1.01 with standard error of 0.5. However, the results did not support a significant correlation between two random components $\mathbf{u}_1$ and $\mathbf{u}_2$ with REML estimated correlation value of $\hat{\rho}$ = -0.11. The REML estimates $\hat{\varphi}_1$ and $\hat{\varphi}_2$ show that variation between customers increases from first question to the second question. The predicted values of the second random component ($\hat{u}_{2i}$) were used to identify customers who showed greater changes (increase or decrease over questions) and 30 customers were identified for further study.

Let $\boldsymbol{\beta}$ be the vector of $\boldsymbol{\beta}_s$, $s$ = 1,2,…13. A Wald test which compares $\hat{\boldsymbol{\beta}}'[\text{var}(\hat{\boldsymbol{\beta}})]^{-1}\hat{\boldsymbol{\beta}}$ = 17.43 (under Model 4) to a $\chi^2_{13}$ value is not significant. Thus, the results do not support a significant difference between two methods of collecting data (face to face and mailing) on overall. In contrast model 1 shows marginally significant effect of the method of collecting data with $\hat{\boldsymbol{\beta}}'[\text{var}(\hat{\boldsymbol{\beta}})]^{-1}\hat{\boldsymbol{\beta}} = 20.5$.

The sub-activities have calculated Chi-Square of 16.5 (under Model 4) which is greater than $\chi^2_8$ at 5% significant level. This indicates that there is a significant variation between sub-activities on the product and services by MSC. The REML estimated values are negative indicating that there is evidence that sub-activities are happy with product and services by MSC.

The threshold parameters do also change significantly over questions. Figure 1 shows the estimated threshold parameters over 13 questions. Therefore, a suitable model to analysis customer satisfaction data is the question dependent random components threshold model. Question independent models may lead to a wrong decision about factors in the model.

Table 2 shows the predicted probability and cumulative probability under Model 4. The questions 9 (the manner of claim back dealing with nonconforming product with customer demands) and 13 (technical support) by MSC have C2 (predicted cumulative probabilities for response category good (2)) of 0.28 and 0.43 respectively. These are small and showing that customers are not happy with the manner of claim back and technical support. Customers are very happy with the product quality from the hot and cold rolling sections of MSC with C2 of 0.76 and 0.77 respectively. The results (not given here) support the previous conclusions that there is no significant difference in cumulative probabilities for response category good (2) between two methods of collecting data.

Table 3 shows the predicted cumulative probabilities for response category good (2) for sub-activities. The results show that the sub-activities are also not happy with the manner of claim back by MSC with C2 smaller than 50%. The pipe and profile and tank, cylinder and container makers are 50 - 50 with technical support by MSC. The sub-activities home accessories makers, steel industries and other user of the steel sheets have C2 less than 50% for the first question (familiarity with MSC and its products). The results also show that all 8 sub-activities are happy with product quality from the hot and cold rolling sections of the MSC. The results (not reported in here) show that almost 10% and 21% of customers are happy (with C2 greater than 60%) by the manner of claim back and technical support of the MSC. Figures 2 and 3 show the predicted cumulative probabilities for response category good (y = 2) for all 202 customers. The results indicate that almost 44% of the customers are not happy with questions 1 (familiarity with the MSC), 2 (dimensions tolerance by the hot rolling section) and 7 (dimensions tolerance by the pickling section) with C2 less than 0.5.

## 6. Discussion

Two random component threshold models are fitted to the customer's opinion data. The threshold parameters are independent of the question in the first model whereas they are varying by questions in the second model. The models are called *question independent* and *question dependent random component threshold models*. The results of the application indicate that the threshold parameters do change significantly over questions. Therefore, a suitable model for analysing customer satisfaction data is the question dependent random components threshold model. Question independent models are led to a wrong decision about factors in the model. The results of question independent random component threshold models indicate that there is a significant difference between two types of collecting data whereas question dependent random component threshold models do not support such a conclusion.

We have also introduced a second random component into the linear predicttor $\boldsymbol{\eta}$ to allow possible change in variance and pattern of association on the second and following questions. The results of application show that the customer effect changes significantly over questions. However, there is no evidence of dependence between two random components. Models lead to a wrong decision about factors if the significant random effects and dependence between them are not included. The estimation approach also provides predicted values of random effects that are very useful in identifying happy and unhappy customers.

## Appendix A

Let y be the observation vector of the ordinal response variable Y. Let $\boldsymbol{\eta}$ be the corresponding vector of linear combination of explanatory variables and random components, given by $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$; $\mathbf{u}$ is the vector of random components and $\mathbf{X}$ and $\mathbf{Z}$ are regression and incidence matrices. Let $l_1$ be the log-likelihood function of y observations conditional on the value of the random component vector $\mathbf{u}$ and let $l_2$ be the logarithm of the probability density function of $\mathbf{u}$. The functions $l_1$ and $l_2$ are

$$l_1 = \sum_{i=1}^{202}\sum_{s=1}^{13} \delta_{is} \ln \Delta_{is}$$
$$l_2 = -(1/2)[const. + \ln|\mathbf{A}| + \mathbf{u}'\mathbf{A}^{-1}\mathbf{u}]$$

where

$$\Delta_{is} = \mathrm{G}(\theta_{y_{is}} - \eta_{is}) - \mathrm{G}(\theta_{y_{is}-1} - \eta_{is}) \quad \text{and}$$

$$\Delta_{is} = \mathrm{G}(\theta_{sy_{is}} - \eta_{is}) - \mathrm{G}(\theta_{s(y_{is}-1)} - \eta_{is})$$

under models (3.1) and(3.2) respectively. The $\delta_{is}$ is one if $i$th customer uses product or service of the $s$th question, zero otherwise. Penalised likelyhood (PL) estimates $\tilde{\boldsymbol{\theta}}$, $\tilde{\boldsymbol{\beta}}$ and $\tilde{\mathbf{u}}$ are obtained by maximising $l = l_1 + l_2$ with respect to $\boldsymbol{\theta}$, $\boldsymbol{\beta}$ and $\mathbf{u}$ respectively. These estimates are then used as an initial step in finding ML and REML estimates of $\varphi_j$ and $\varphi_3$ via Anderson (1973) and Henderson (1973) algorithm. The iterative procedure used to obtain the ML and REML estimators and their approximate variance-covariance matrices can be specified as follows:

(a) Let $\mathbf{I}_\theta$ to be an identity matrix of dimension equal to the number of the rows in θ, $\overset{*}{\mathbf{X}} = \begin{bmatrix} \mathbf{I}_\theta & 0 & 0 \\ 0 & \mathbf{X} & \mathbf{Z} \end{bmatrix}$ and $k$ =0. Starting from initial values $\boldsymbol{\theta}_0$, $\boldsymbol{\beta}_0$ , $\mathbf{u}_0$, $\varphi_{j0}$ and $\varphi_{30}$ (hence $\mathbf{A}_0$), the estimating equations are:

$$(3) \qquad \begin{bmatrix} \boldsymbol{\theta}_{k+1} \\ \boldsymbol{\beta}_{k+1} \\ \mathbf{u}_{k+1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\theta}_{k} \\ \boldsymbol{\beta}_{k} \\ \mathbf{u}_{k} \end{bmatrix} + \mathbf{V}^{-1}\overset{*}{\mathbf{X}}{}' \begin{bmatrix} \partial l_1 / \partial \boldsymbol{\theta}_k \\ \partial l_1 / \partial \boldsymbol{\eta}_k \end{bmatrix} - \mathbf{V}^{-1} \begin{bmatrix} 0 \\ 0 \\ \mathbf{A}_0^{-1}\mathbf{u}_k \end{bmatrix}$$

where $\mathbf{V} = \overset{*}{\mathbf{X}}{}' \left( - \begin{bmatrix} \partial^2 l_1 / \partial \boldsymbol{\theta}_k \partial \boldsymbol{\theta}'_k & \partial^2 l_1 / \partial \boldsymbol{\theta}_k \partial \boldsymbol{\eta}'_k \\ \partial^2 l_1 / \partial \boldsymbol{\eta}_k \partial \boldsymbol{\theta}'_k & \partial^2 l_1 / \partial \boldsymbol{\eta}_k \partial \boldsymbol{\eta}'_k \end{bmatrix} \right) \overset{*}{\mathbf{X}} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \mathbf{A}_0^{-1} \end{bmatrix}$,

$\partial l_1 / \partial \boldsymbol{\theta}_k$ , $\partial l_1 / \partial \boldsymbol{\eta}_k$ , $\partial^2 l_1 / \partial \boldsymbol{\theta}_k \partial \boldsymbol{\theta}'_k$, $\partial^2 l_1 / \partial \boldsymbol{\theta}_k \partial \boldsymbol{\eta}'_k$ and $\partial^2 l_1 / \partial \boldsymbol{\eta}_k \partial \boldsymbol{\eta}'_k$ are evaluated at initial estimates of the parameters.

(b) Once iterations of (3) have converged to $\tilde{\boldsymbol{\beta}}$ and $\tilde{\mathbf{u}}$, let $\tilde{\boldsymbol{\eta}} = \mathbf{X}\tilde{\boldsymbol{\beta}} + \mathbf{Z}\tilde{\mathbf{u}}$, $\mathbf{B} = -\partial^2 l_1 / \partial \tilde{\boldsymbol{\eta}} \partial \tilde{\boldsymbol{\eta}}'$, $\mathbf{T}^* = [\mathbf{T}^*_{jj'}] = [\mathbf{A}_0^{-1} + \mathbf{Z}^{\mathrm{T}}\mathbf{B}\mathbf{Z}]^{-1}$, $a_{jj'} = \mathrm{tr}(\mathbf{T}^*_{jj'} + \mathbf{T}^*_{j'j}) + 2\tilde{\mathbf{u}}'_j\tilde{\mathbf{u}}_{j'}$ and $a_j = \mathrm{tr}(\mathbf{T}^*_{jj}) + \tilde{\mathbf{u}}'_j\tilde{\mathbf{u}}_j$ for $j, j' = 1, 2$. The ML estimates of $\varphi_j$ and $\varphi_3$ are

$$(4) \qquad \begin{aligned} \hat{\varphi}_{j(ML)} &= N^{-1}[a_j + (\varphi_{30}/\varphi_{j'})^2 a_{j'} + (N\varphi_{30}^2/\varphi_{j'}) - (\varphi_{30}/\varphi_{j'})a_{jj'}] \\ \hat{\varphi}_{3(ML)} &= \varphi_{10}\varphi_{20}a_{12}[2N\varphi_{30}^2 - a_{12}\varphi_{30} + \varphi_{20}a_1 + \varphi_{10}a_2 - 2\varphi_{10}\varphi_{20}N]^{-1} \end{aligned} \quad \text{for } j = 1, 2.$$

The preceding two steps are then repeated, with $k$ =0, and initial values set to $\tilde{\boldsymbol{\theta}}$, $\tilde{\boldsymbol{\beta}}$, $\tilde{\mathbf{u}}$, $\hat{\varphi}_{j(ML)}$ and $\hat{\varphi}_{3(ML)}$.

Once this iterative process has converged, the asymptotic variance-covariance matrix for the ML estimators $\varphi_{ML} = (\hat{\varphi}_{1(ML)}, \hat{\varphi}_{2(ML)}, \hat{\varphi}_{3(ML)})'$ is

(5) $\quad \mathrm{var}(\varphi_{ML}) = 2\left[r_{1jj'} + r_{2jj'} - 2r_{3jj'}\right]^{-1}$

where $r_{1jj'} = tr[\mathbf{A}(\partial\mathbf{A}/\partial\varphi_j)\mathbf{A}(\partial\mathbf{A}/\partial\varphi_{j'})]$, $r_{2jj'} = tr[\mathbf{T}^*(\partial\mathbf{A}/\partial\varphi_j)\mathbf{T}^*(\partial\mathbf{A}/\partial\varphi_{j'})]$ and $r_{3jj'} = tr[\mathbf{A}(\partial\mathbf{A}/\partial\varphi_j)\mathbf{T}^*(\partial\mathbf{A}/\partial\varphi_{j'})]$.

Let $\begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} & \mathbf{V}_{13} \\ \mathbf{V}_{21} & \mathbf{V}_{22} & \mathbf{V}_{23} \\ \mathbf{V}_{31} & \mathbf{V}_{32} & \mathbf{V}_{33} \end{bmatrix}$ denote the partitions of the matrix $\mathbf{V}$ corresponding to the dimensions of $\boldsymbol{\theta}$, $\boldsymbol{\beta}$ and $\mathbf{u}$, similarly let $\mathbf{V}^{-1} = \begin{bmatrix} \mathbf{T}_{11} & \mathbf{T}_{12} & \mathbf{T}_{13} \\ \mathbf{T}_{21} & \mathbf{T}_{22} & \mathbf{T}_{23} \\ \mathbf{T}_{31} & \mathbf{T}_{32} & \mathbf{T}_{33} \end{bmatrix}$.

Replacing $\mathbf{T}^*$ by $\mathbf{T}_{33}$ in (4) and (5) yields REML estimates $\hat{\varphi}_{j(REML)}$ and $\hat{\varphi}_{3(REML)}$ and their corresponding asymptotic variance-covariance matrix.

## References


Anderson, T.W. (1973), "Asymptotically efficient estimation of covariance matrices with linear structure," *Annals of Statistics*, **1** 135 - 141.

Anderson, J. A. and Philips, P. R. (1981), "Regression, discrimination and measurement models ordered categorical variables," *Journal of Applied Statistics*, **30**, 22 - 31.

Breslow, N.E. and Clayton, D.G. (1993), "Approximate inference in generalised linear mixed models," *Journal of the American Statistical Association*, **88**, 9 - 25.

Crouchley, R. (1995), "A random effects model for ordered categorycal data," *Journal of the American Statistical Association* **90,** 489 - 498.

Gianola, D. and Foulley, I.J. (1983), "Sire evaluation for ordered categorical data with a threshold model," *Genetics Selection Evaluation*.

Harville, D.A. and Mee, R.W. (1984), "A mixed model procedure for analysing ordered categorical data," *Biometrics* **40**, 393 - 944.

Henderson, C.R.(1973b), Maximum likelihood estimation of variance components, *unpublished manuscript*.

Jansen, J. (1990), "On the statistical analysis of ordinal data when extra variation is present," *Applied Statistics* **39**, 75 - 84.

Jansen, J. (1992), "Statistical analysis of ordinal data from experiments with nested errors," *Computational Statistics & Data Analysis* **13**, 319 - 330.

McCullagh, P. and Nelder, J.A. (1989), *Generalized linear models.* Second Ed. London: Chapman and Hall.

McGilchrist, C.A. (1994), "Estimation in generalized mixed models," *Journal of the Royal Statistical Society, Series* **B 56,** 61 - 69.

Nelder, J.A. and Lee, Y. (1996), "Hierarchical generalized linear models," *Journal of the Royal Statistical Society, Series* **B 58,** 619 - 678.

Lee, Y. and Nelder, J.A. (2001a), "Hierarchical generalized linear models: A synthesis of generalised linear models, random effect model and structured dispersions," *Biometrika*, **88**, 987 - 1006.

Lee, Y. and Nelder, J.A. (2001b), "Modelling and analysing correlated non-normal data," *Statistical Modelling*, **1**, 3 - 16.

Saei, A. and McGilchrist, C.A. (1996), "Random component threshold models," *Journal of Agricultural, Biological and Environmental Statistics*, **1**, 288 - 296.

Saei, A. and McGilchrist, C. (1998), "Longitudinal threshold models with random components," *The Statistician , Series* **D 47**, 365 - 375.

Schall, R (1991), "Estimation in generalised linear models with random effects," *Biometrika* **78**, 719 - 727.

Ten Have T. (1996), "A mixed effects model for multivariate ordinal response data including correlated discrete failure time with ordinal responses," *Biometrics*, **53**, 473 - 491.

Wolfinger, R. (1993), "Laplace's approximation for nonlinear mixed models," *Biometrika* **80,** 791 - 796.

Zhaorong, J., McGilchrist, C.A. and Jorgensen, M.A. (1992), "Mixed model discrete regression," *Biometrical Journal* **34**, 691 - 700.

**Table 1: REML** estimate of parameters (Est), standard error (SE) and *Z*-v = (Est/SE).

| Model | Model 1 | | | Model 2 | | | Model 3 | | | Model 4 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Est | SE | Z-v | Est | SE | Z-v | Est | SE | Z-v | Est | SE | Z-v |
| $\varphi_1$ | | | | 1.49 | 0.2 | 7.37 | 0.83 | 0.24 | 3.54 | 0.94 | 0.24 | 3.54 |
| $\varphi_2$ | | | | | | | 0.92 | 0.27 | 3.44 | 1.02 | 0.27 | 3.44 |
| $\varphi_3$ | | | | | | | | | | -0.1 | 0.41 | -0.26 |
| Mail 1 | 0.31 | 0.28 | 1.11 | 0.26 | 0.34 | 0.78 | 0.27 | 0.31 | 0.87 | 0.27 | 0.31 | 0.86 |
| Mail 2 | 0.51 | 0.32 | 1.6 | 0.52 | 0.39 | 1.36 | 0.5 | 0.4 | 1.27 | 0.5 | 0.4 | 1.27 |
| Mail 3 | -0.46 | 0.31 | -1.5 | -0.7 | 0.37 | -1.89 | -0.75 | 0.38 | -1.98 | -0.75 | 0.38 | -1.98 |
| Mail 4 | 0.54 | 0.41 | 1.31 | 0.23 | 0.47 | 0.49 | 0.19 | 0.48 | 0.4 | 0.19 | 0.48 | 0.4 |
| Mail 5 | 0.43 | 0.33 | 1.31 | 0.4 | 0.38 | 1.05 | 0.39 | 0.39 | 1 | 0.39 | 0.39 | 1.01 |
| Mail 6 | 0.31 | 0.31 | 1 | 0.19 | 0.37 | 0.51 | 0.15 | 0.37 | 0.4 | 0.15 | 0.37 | 0.4 |
| Mail 7 | 0.77 | 0.44 | 1.75 | 0.53 | 0.5 | 1.08 | 0.5 | 0.51 | 0.98 | 0.5 | 0.51 | 0.98 |
| Mail 8 | -0.21 | 0.3 | -0.69 | -0.22 | 0.36 | -0.6 | -0.25 | 0.37 | -0.66 | -0.25 | 0.37 | -0.66 |
| Mail 9 | -0.38 | 0.37 | -1.02 | -0.48 | 0.42 | -1.14 | -0.52 | 0.43 | -1.2 | -0.52 | 0.43 | -1.19 |
| Mail 10 | 0.16 | 0.34 | 0.48 | 0.02 | 0.4 | 0.05 | 0.01 | 0.41 | 0.02 | 0.01 | 0.41 | 0.02 |
| Mail 11 | 0.05 | 0.33 | 0.16 | -0.13 | 0.39 | -0.32 | -0.18 | 0.4 | -0.45 | -0.18 | 0.4 | -0.44 |
| Mail 12 | 1.07 | 0.49 | 2.21 | 0.9 | 0.55 | 1.65 | 0.88 | 0.56 | 1.58 | 0.88 | 0.56 | 1.59 |
| Mail 13 | -0.21 | 0.36 | -0.58 | -0.42 | 0.42 | -1 | -0.45 | 0.43 | -1.03 | -0.44 | 0.43 | -1.03 |
| PPM | -1.05 | 0.27 | -3.84 | -1.48 | 0.59 | -2.5 | -1.17 | 0.59 | -1.99 | -1.18 | 0.59 | -2.01 |
| TCCM | -0.93 | 0.32 | -2.9 | -1.15 | 0.66 | -1.76 | -0.82 | 0.64 | -1.28 | -0.84 | 0.64 | -1.3 |
| HAM | -0.01 | 0.27 | -0.02 | -0.28 | 0.57 | -0.49 | -0.08 | 0.56 | -0.15 | -0.09 | 0.56 | -0.16 |
| SI | -0.47 | 0.27 | -1.73 | -0.83 | 0.58 | -1.45 | -0.38 | 0.57 | -0.67 | -0.4 | 0.57 | -0.7 |
| OUSS | -0.04 | 0.3 | -0.13 | -0.38 | 0.64 | -0.59 | -0.08 | 0.63 | -0.13 | -0.09 | 0.63 | -0.14 |
| HSE | -0.9 | 0.27 | -3.39 | -1.17 | 0.56 | -2.1 | -0.91 | 0.55 | -1.66 | -0.92 | 0.55 | -1.68 |
| SSCU | -1 | 0.29 | -3.43 | -1.38 | 0.64 | -2.15 | -0.95 | 0.63 | -1.51 | -0.97 | 0.63 | -1.54 |
| MCM | -0.39 | 0.28 | -1.42 | -0.54 | 0.59 | -0.93 | -0.34 | 0.58 | -0.59 | -0.33 | 0.58 | -0.6 |

- Model 1 = Fixed effect threshold model; Model 2 = A one random component threshold model; Model 3 = A two independent random components threshold model; Model 4 = A two random components threshold model
- Mail *i* is the mail effect for the question *i*, *i* = 1, 2,…, 13
- Water, oil and gas pipe makers and face – face effects are fixed at zero for identifiably
- PPM = pipe and profile makers, TCCM = tank, cylinder and container makers, HAM = home accessories makers, SI = steel industries, OUSS = other user of steel, sheet, HSE = heavy steel equipment, SSCU = steel sheet cooperation users, MCM = motor car makers.

**Table 2: REML** predicted probability and cumulative probability of the response categories.

| | Category | | | | Cumulative | | |
|---|---|---|---|---|---|---|---|
| Question | 1 | 2 | 3 | 4 | C1 | C2 | C3 |
| 1 | 0.09 | 0.43 | 0.38 | 0.1 | 0.09 | 0.52 | 0.9 |
| 2 | 0.3 | 0.46 | 0.22 | 0.02 | 0.3 | 0.76 | 0.98 |
| 3 | 0.3 | 0.47 | 0.2 | 0.03 | 0.3 | 0.77 | 0.97 |
| 4 | 0.19 | 0.43 | 0.25 | 0.13 | 0.19 | 0.62 | 0.87 |
| 5 | 0.21 | 0.35 | 0.31 | 0.12 | 0.21 | 0.56 | 0.87 |
| 6 | 0.2 | 0.44 | 0.26 | 0.09 | 0.2 | 0.64 | 0.9 |
| 7 | 0.14 | 0.42 | 0.33 | 0.11 | 0.14 | 0.56 | 0.89 |
| 8 | 0.18 | 0.44 | 0.27 | 0.12 | 0.18 | 0.62 | 0.89 |
| 9 | 0.06 | 0.23 | 0.36 | 0.35 | 0.06 | 0.29 | 0.65 |
| 10 | 0.1 | 0.54 | 0.31 | 0.05 | 0.1 | 0.64 | 0.95 |
| 11 | 0.09 | 0.55 | 0.34 | 0.02 | 0.09 | 0.64 | 0.98 |
| 12 | 0.08 | 0.58 | 0.33 | 0.01 | 0.08 | 0.66 | 0.99 |
| 13 | 0.11 | 0.32 | 0.27 | 0.31 | 0.11 | 0.43 | 0.7 |

- Category: 1 = very good, 2 = good, 3 = not bad and 4 = bad
- Cumulative: C1 = $P(Y \leq 1)$, C2 = $P(Y \leq 2)$ and C3 = $P(Y \leq 3)$.

**Table 3**: REML predicted cumulative probability of the second category, (good), for 8 sub-activities.

| | Sub-activity | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Question | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 1 | 0.63 | 0.59 | 0.43 | 0.48 | 0.42 | 0.6 | 0.6 | 0.48 |
| 2 | 0.84 | 0.79 | 0.71 | 0.77 | 0.69 | 0.8 | 0.81 | 0.74 |
| 3 | 0.86 | 0.8 | 0.72 | 0.81 | 0.72 | 0.74 | 0.89 | 0.76 |
| 4 | 0.71 | 0.64 | 0.57 | 0.64 | 0.56 | 0.6 | 0.7 | 0.6 |
| 5 | 0.67 | 0.58 | 0.49 | 0.57 | 0.49 | 0.62 | 0.63 | 0.53 |
| 6 | 0.74 | 0.7 | 0.58 | 0.68 | 0.57 | 0.63 | 0.75 | 0.63 |
| 7 | 0.64 | 0.6 | 0.53 | 0.58 | 0.51 | 0.55 | 0.64 | 0.54 |
| 8 | 0.71 | 0.67 | 0.54 | 0.63 | 0.52 | 0.64 | 0.72 | 0.57 |
| 9 | 0.35 | 0.34 | 0.21 | 0.3 | 0.25 | 0.32 | 0.42 | 0.24 |
| 10 | 0.74 | 0.64 | 0.55 | 0.63 | 0.55 | 0.68 | 0.72 | 0.58 |
| 11 | 0.76 | 0.68 | 0.58 | 0.67 | 0.57 | 0.62 | 0.75 | 0.62 |
| 12 | 0.71 | 0.72 | 0.64 | 0.7 | 0.58 | 0.65 | 0.72 | 0.65 |
| 13 | 0.5 | 0.48 | 0.37 | 0.44 | 0.38 | 0.45 | 0.43 | 0.39 |

- Sub-activity: 1 = pipe and profile makers, 2 = tank, cylinder and container makers, 3 = home, accessories makers, 4 = steel industries, 5 = other user of steel sheet, 6 = heavy steel, equipment, 7 = steel sheet cooperation users, 8 = motor car makers.

**Figure 1: REML** estimates of the threshold parameters on each 13 questions.

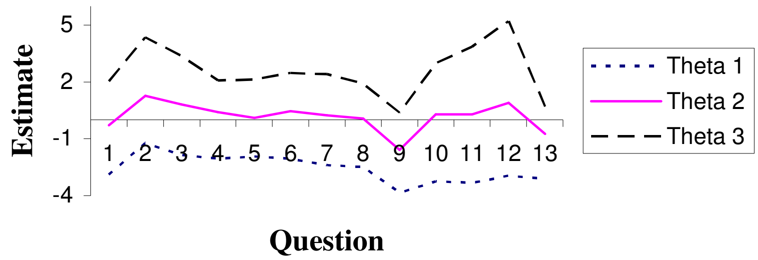


**Figure 2: REML** predicted cumulative probability of the response category 2 ( P(Y≤2)) for the question 9 (the manner of claim by MSC).

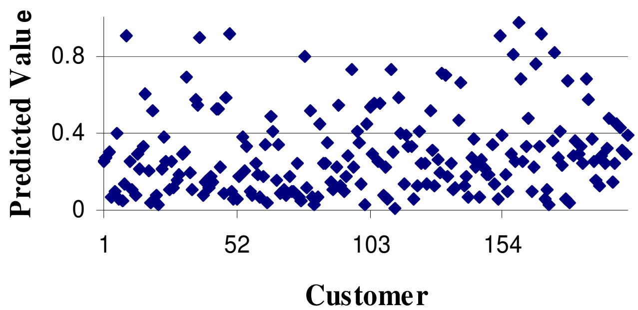


**Figure 3: REML** predicted cumulative probability of the response category 2 ( P(Y≤2)) for the question 13 (Technical support by MSC).