GLMM Based Segmentation of Czech Households Using the EU-SILC Database

Jan Vávra, Arnošt Komárek[1]

Abstract. The EU-SILC database contains annually gathered rotating-panel data on a household level covering indicators of monetary poverty, severe material deprivation or low work household intensity. Data are obtained via questionnaires leading to outcome variables of diverse nature: numeric, binary, ordinal or general categorical. In our previous contribution to MME 2020 we presented a clustering method for such a type of data. The used thresholding approach of latent numeric counterparts of binary and ordinal outcomes suffered from slow convergence and unclear interpretation of resulting estimates. Hence we propose an alternative approach which again exploits a Bayesian variant of the Model Based Clustering (MBC). Nevertheless, the underlying models are all of a Generalized Linear Mixed Model (GLMM) nature: (proportional odds) logit model for (ordinal) or binary indicators, multinomial logit model for general categorical outcomes and a standard linear mixed model for numeric outcome. Czech households interviewed within the EU-SILC project between 2005 and 2018 are then divided into several groups of similar evolution of income, housing costs, self-evaluations and other indicators.

Keywords: Multivariate panel data, Mixed type outcome, GLMM, Model based clustering, Classification

JEL Classification: C33, C38
AMS Classification: 62H30

1 Introduction

Throughout the EU states the poverty and social exclusion is measured using indicators of monetary poverty, severe material deprivation and very low work household intensity. Relevant data are gathered within *The European Union Statistics on Income and Living Conditions* project (EU-SILC, https://ec.europa.eu/eurostat/web/microdata/european-union-statistics-on-income-and-living-conditions. This is an instrument with the goal to collect timely and comparable cross-sectional and *longitudinal multidimensional* microdata on income, poverty, social exclusion and living conditions. Data are obtained via questionnaires leading to outcome variables of diverse nature: *numeric* (e.g., income), *binary* (e.g., affordability of paying unexpected expenses), *ordinal* (e.g., level of ability to make ends meet) or *categorical* (e.g., ownership of a car/computer). It is our primary aim to use such longitudinally gathered outcomes towards segmentation of households according to typical patterns of their temporal evolution.

In our previous contribution [6] to MME 2020 we proposed a statistical model capable of joint modelling of longitudinal outcomes of diverse nature (*numeric, binary, ordinal*). We improve it by replacing thresholding of latent numeric outcomes with generalized linear mixed models (GLMM, [4]) appropriate to the type of the modelled outcomes. This change also allows us to extend the model for completely general categorical outcomes using multinomial logistic regression with random effects. This is a topic of Section 2. Consequently, we apply model based clustering (MBC, [3]) procedure to perform unsupervised classification of study units (households) into groups whose characteristics are not known in advance. This part of methodology is described in Section 3. We also dedicate one additional Section 4 to explain in more details, how do we use Markov chain Monte Carlo (MCMC, [2]) methods to estimate the model in Bayesian setting. The final Section 5 describes the use of this methodology on the Czech subset of the EU-SILC dataset. The paper is finalized by conclusions in Section 6.

2 Joint model for mixed type panel data

In general, we have data on n units/panel members (e.g., households) at our disposal containing $R \geq 1$ longitudinally gathered outcomes (e.g., income, affordability of week holiday and level of a financial burden of housing). Let $\mathbf{Y}_i = \left(\mathbf{Y}_{i,1}^\top, \ldots, \mathbf{Y}_{i,R}^\top\right)^\top$ stand for a vector consisting of all the values $\mathbf{Y}_{i,r} = (Y_{i,r,1}, \ldots, Y_{i,r,n_i})^\top$ of the rth outcome ($r = 1, \ldots, R$) of the ith unit ($i = 1, \ldots, n$) obtained at n_i occasions. Let C_i stand for available covariates (the

[1] Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic, Department of Probability and Mathematical Statistics, vavraj@karlin.mff.cuni.cz

times of measurements, possibly other explanatory variables) of i-th unit. Finally, let $h(\mathbf{y}_i;\, C_i,\, \boldsymbol{\zeta})$ represent the assumed distribution of the outcome vector $\mathbf{Y}_i$ which possibly depends on the covariates C_i and also on a vector $\boldsymbol{\zeta}$ of unknown parameters.

This setting corresponds to the one considered in [6], however the assumed distribution h will be different. Previously, we assumed linear mixed model (LMM, [5]) for numeric outcomes and for latent numeric counterparts of binary or ordinal outcomes. The observed categorical outcomes were then considered to be result of segmentation into intervals by unknown set of thresholds. However, the estimation of these parameters exhibited very slow convergence to posterior distribution (see Section 4). Moreover, estimated coefficients were tied to the dimensionless latent outcome and not directly to the observed ordered categories. For this reasons we decided to leave thresholding approach in favour of generalized linear mixed models (GLMM, [4]), which still allows us to join models for different outcomes through general combined distribution of random effects.

For each outcome r we consider a predictor $\eta_{i,r,j} = \eta^F_{i,r,j} + \eta^R_{i,r,j}$ for jth observation of ith unit that consists of

- fixed part $\eta^F_{i,r,j} = \mathbf{X}^\top_{i,r,j}\boldsymbol{\beta}_r$, where $\boldsymbol{\beta}_r$ are unknown fixed effects of regressors $\mathbf{X}_{i,r,j}$ created from covariates $C_{i,j}$.
- random part $\eta^R_{i,r,j} = \mathbf{X}^\top_{i,r,j}\mathbf{B}_{i,r}$, where $\mathbf{B}_{i,r}$ are unit-specific unknown random effects of regressors $\mathbf{Z}_{i,r,j}$ created from covariates $C_{i,j}$.

The distribution of observed outcomes $Y_{i,r,j}$ is then supposed to depend on this predictor $\eta_{i,r,j}$ depending on what nature the outcome is. Let us for the moment drop the indices i, r, j.

- *Numeric* outcome Y is assumed to follow normal distribution with mean value η and precision parameter τ, which is a reciprocal value to standard variance parameter. This distribution can be summarized via logarithm of probability density function (log-pdf) $\ell^{\mathsf{N}}(Y|\eta,\tau) = -\frac{1}{2}\log(2\pi) + \frac{1}{2}\log\tau - \frac{1}{2}\tau\,(Y-\eta)^2$.
- *Binary* outcome $Y \in \{0,1\}$ is assumed to follow logistic regression model, which parametrizes probability of $Y = 1$ by inverse of *logit* function: $\mathsf{P}\,[Y = 1|\eta] = \mathrm{logit}^{-1}(\eta) = e^\eta/(1+e^\eta)$. The corresponding log-pdf is then of form $\ell^{\mathsf{B}}(Y|\eta) = Y\cdot\eta - \log(1+e^\eta)$.
- *Ordinal* outcome $Y \in \{1,\ldots,K\}$ of K ordered levels is assumed to follow ordinal logistic regression (ORL), where the probability of attaining level greater than $k \in \{1,\ldots,K\}$ is parametrized by $p_k := \mathsf{P}\,[Y > k|\eta,\mathbf{c}] = \mathrm{logit}^{-1}(\eta - c_k)$, where $\mathbf{c} = (c_1,\ldots,c_{K-1})$ is a set of ordered thresholds $-\infty = c_0 < c_1 < \cdots < c_{K-1} < c_K = \infty$ that play a role of intercept. For identifiability purposes, the predictor η must not contain a fixed intercept term (or fix corresponding β parameter to zero). We extend our notation to $p_0 = 1$ and $p_K = 0$. The probability q_k of attaining level k can then be expressed as $q_k = p_{k-1} - p_k$, which finally yields the corresponding log-pdf $\ell^{\mathsf{O}}(Y|\eta,\mathbf{c}) = \log q_Y = \log(p_{Y-1} - p_Y) = \mathrm{logit}^{-1}(\eta - c_{Y-1}) - \mathrm{logit}^{-1}(\eta - c_Y)$.
- *General categorical* outcome $Y \in \{1,\ldots,K\}$ of K unordered levels is the newest addition to the model, for which we decided to use multinomial logistic regression (MLR) model that parametrizes probability of attaining level $k \in \{1,\ldots,K\}$ by K different predictors η_k that share the same structure, however, they use different set of fixed effects $\boldsymbol{\beta}_{r,k}$. For simplicity of the model we use only one set of random effects $B_{i,r}$ creating only one η^{R} for all predictors η_k. In order to identify these effects, it is necessary to fix them to zero for one k, let us use $\eta_K = 0$. Then the probability that Y falls into category $k \in \{1,\ldots,K\}$ is kth element of so called softmax function $\mathsf{P}\,[Y = k|\eta_1,\ldots,\eta_{K-1}] = \mathrm{softmax}_k(\boldsymbol{\eta}) = e^{\eta_k}\Big/\left(1 + \sum_{k'=1}^{K-1} e^{\eta_{k'}}\right)$, where $\boldsymbol{\eta} = (\eta_1,\ldots,\eta_K)$ is vector of all predictors corresponding to categorical outcome Y. The resulting log-pdf is then of the form $\ell^{\mathsf{C}}(Y|\boldsymbol{\eta}) = \eta_Y - \log\left(1 + \sum_{k=1}^{K-1} e^{\eta_k}\right)$.

Note that under $K = 2$ both ordinal and categorical setting reduce to the logistic regression assumed for binary outcomes. Such an ordinal outcome would require lone threshold c_1 that would correspond to negative intercept term in logistic regression since $q_1 = 1 - p_1$ and $q_2 = p_1$. On the other hand, a categorical outcome would then have one actual predictor $\eta = \eta_1$ and $\eta_2 = 0$ which trivially means that the first element of softmax coincides with logit^{-1} in logistic regression. To avoid such special cases we limit ourselves to $K \geq 3$ when using ORL or MLR, and hence any categorical outcome of $K = 2$ levels will be treated by logistic regression directly.

Now, when supposed distributions for observed outcomes are set, we discuss the other object of randomness - random effects. Let us for ith unit denote $\mathbf{B}_i$ a vector of all random effects $\mathbf{B}_{i,r}$. For clarity, we will always consider outcomes to be ordered by their type - first numeric, then binary, ordinal and categorical as last and in this way we also order elements of vector of random effects $\mathbf{B}_i$. By the assumption that $\mathbf{B}_i \overset{\text{iid}}{\sim} \mathsf{N}(\mathbf{0},\boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} > 0$ is unknown completely general covariance matrix, we incorporate possible dependencies among outcomes into our model. Because even when the conditional distributions of different outcomes given random effects are independent, the unconditional distributions become dependent. Such distribution is possible to derive when only numeric outcomes are considered, however, once we add any categorical outcome modelled by GLMM it becomes much more challenging task. In fact, the overall pdf h for $\mathbf{Y}_i$ in general takes the form of

$$h(\mathbf{y}_i;\, C_i,\, \zeta) = \int \prod_{r=1}^{R} \prod_{j=1}^{n_i} \exp\left\{\ell^{\mathsf{type}(r)}\left(Y_{i,r,j}|C_{i,j}, \mathbf{B}_{i,r}, \zeta_r\right)\right\} \cdot (2\pi)^{-\frac{\dim \mathbf{B}_i}{2}} |\mathbf{\Sigma}|^{-1} \exp\left\{-\frac{1}{2}\mathbf{B}_i^\top \mathbf{\Sigma}^{-1}\mathbf{B}_i\right\} \mathrm{d}\mathbf{B}_i, \tag{1}$$

where $\mathsf{type}(r) \in \{\mathsf{N}, \mathsf{B}, \mathsf{O}, \mathsf{C}\}$ denotes the type of rth outcome and ζ consists of all ζ_r that cover all unknown parameters tied with rth outcome. We propose to solve this integration by multivariate version of adaptive Gauss-Hermite quadrature using Laplace approximation (Breslow and Clayton, [1]). Our MCMC based estimation completely avoids the necessity of performing this integration. However, it is still needed for detailed exploration of posterior distribution of several useful characteristics.

3 Model based clustering

In this section we discuss the way we can distinguish different patterns in joint model for mixed type panel data. We are given with the number of groups G into which we intend to classify the units in advance and $G \geq 2$. The classification proceeds by using the model outlined in Section 2 within the Bayesian model based clustering procedure (MBC, [3]). It is assumed that the overall model, f, is given as a finite mixture of certain group-specific models f_g, $g = 1, \ldots, G$. That is, $f(\mathbf{y}_i;\, C_i,\, \boldsymbol{\theta}) = \sum_{g=1}^{G} w_g\, f_g(\mathbf{y}_i;\, C_i,\, \boldsymbol{\psi},\, \boldsymbol{\psi}^g)$, where $\mathbf{w} = \left(w_1, \ldots, w_G\right)^\top$ are the mixture weights (proportions of the groups in the population), $\boldsymbol{\psi}$ is a vector of unknown parameters common to all groups and $\boldsymbol{\psi}^g$, $g = 1, \ldots, G$, are vectors of group-specific unknown parameters. Hence the vector $\boldsymbol{\theta}$ of all unknown parameters is $\boldsymbol{\theta} \equiv \left\{\mathbf{w},\, \boldsymbol{\psi},\, \boldsymbol{\psi}^1, \ldots, \boldsymbol{\psi}^G\right\}$.

Using the notation from previous section we set the group-specific density f_g to be the density h, however, depending on parameter $\boldsymbol{\zeta}^g$ elements of which ($\boldsymbol{\beta}_r^g$, τ_r^g, $\mathbf{c}_r^g$, $\mathbf{\Sigma}^g$) may (or may not) be group-specific, i.e. different value of the parameter is considered to be in different groups. For example, if we suppose that the groups differ only in the covariate effects, then $\boldsymbol{\psi} = \{\boldsymbol{\tau},\, \mathbf{c},\, \mathbf{\Sigma}\}$ and $\boldsymbol{\psi}^g = \boldsymbol{\beta}^g$.

Further, let $U_i \in \left\{1, \ldots, G\right\}$ denote the unobserved allocation of the ith unit into one of the G groups. As it is usual with the mixture models, the group-specific distribution $f_g(\mathbf{y}_i;\, C_i,\, \boldsymbol{\psi},\, \boldsymbol{\psi}^g)$, $g = 1, \ldots, G$, can be viewed as a conditional distribution of the outcome $\mathbf{Y}_i$ given $U_i = g$ while the mixture weights $\mathbf{w}$ determine the marginal distribution of the allocations, i.e., $\mathsf{P}(U_i = g) = w_g$, $g = 1, \ldots, G$. Classification of the ith unit is then based on suitable estimates of the conditional individual allocation probabilities $p_{i,g}(\boldsymbol{\theta})$, $g = 1, \ldots, G$, calculated by the Bayes rule:

$$p_{i,g}(\boldsymbol{\theta}) \;=\; \mathsf{P}\big(U_i = g \,\big|\, \mathbf{Y}_i = \mathbf{y}_i;\, C_i,\, \boldsymbol{\theta}\big) \;=\; \frac{w_g\, h\left(\mathbf{y}_i;\, C_i,\, \boldsymbol{\psi},\, \boldsymbol{\psi}^g\right)}{f(\mathbf{y}_i;\, C_i,\, \boldsymbol{\theta})}, \tag{2}$$

where $\boldsymbol{\theta} = \{\boldsymbol{\psi}, w_g, \boldsymbol{\psi}^g, g = 1, \ldots, G\}$ is the set of all unknown parameters. Unfortunately, calculation of such probabilities would require immense amount of calculation of the integral (1). However, we can still bypass this issue by taking advantages of Bayesian approach and MCMC estimation, see the end of Section 4.

4 Estimation by MCMC

A Bayesian approach that treats unknown parameters to be random by assigning an uninformative prior is a natural choice for our model. It allows us to consider all random effects $\mathbf{B}_i$ and group allocation indicators U_i as additional unknown parameters of the model. Markov chain with states consisting of all unknown parameters $\boldsymbol{\theta}$, latent $\mathbf{B}_i$ and U_i needs to be constructed in such a way that its limiting distribution corresponds to their posterior distribution, i.e. the distribution of $\boldsymbol{\theta}, \mathbf{B}_i, U_i, i = 1, \ldots, n$ given all available data. Once the chain reaches its stationary distribution (equal to the limiting one) we follow Monte Carlo principles and continue in sampling to obtain a sample of state values on which an estimation of posterior distribution is based.

The question is, how do we construct such a Markov chain. The most straightforward way (taken in [6]) using Gibbs sampling cannot be applied directly, since needed full-conditional distributions of fixed effects $\boldsymbol{\beta}_r^g$, random effects $\mathbf{B}_i$ and ordered intercepts $\mathbf{c}_r^g$ do not fall into known distributional families. Nevertheless, we still keep the main idea of sampling from full-conditional distributions but in cases where the full-conditional distribution is unclear we rather replace it with a Metropolis step that accepts a new sampled value from proposal distribution with a corresponding acceptance probability. This so-called *Metropolis-within-Gibbs* approach satisfies all conditions necessary for MCMC to work [2].

However, there still remains the problem of selecting the proposal distributions in Metropolis steps. A vague choice could lead to poor acceptance probability and inefficient Markov chain. In this regard we adapt the proposal distribution to reflect the true full-conditional distribution behind. By Newton-Raphson method we find the parameter value maximizing full-conditional distribution and evaluate the second order derivatives at this parameter value to obtain a variance matrix for proposal distribution. To be more specific, we transition from current value to the new one where the direction is sampled from centered multivariate normal distribution with such variance matrix possibly modified by a constant to reach better acceptance rate. To reduce autocorrelation (and speed up convergence) we can perform a walk consisting of more than one step. The frequent update of variance matrices of proposal distributions is crucial at the burn-in period of the Markov chain, however, once the stationary distribution is reached it could be updated at lower rate to speed up the sampling.

Finding proposal distributions for fixed effects $\boldsymbol{\beta}_r^g$ is analogous to finding maximum likelihood estimates in GLM models on subset of units in gth cluster. However, under the Bayesian setting we do not maximize only log-pdf, but we also add a logarithm of prior distribution. During Newton-Raphson only fixed part of predictor η^{F} changes while the random part η^{R} is kept the same since we condition by the random effects in this step. The process is very analogous for random effects $\mathbf{B}_i$, however, this time η^{F} remains the same while the random part η^{R} changes. Unlike with the fixed effects $\boldsymbol{\beta}_r^g$, where each was done separately for all $r = 1, \ldots, R$, random effects $\mathbf{B}_i$ have to be sampled jointly across all outcomes due to our common general prior $\mathbf{B}_i | U_i = g \sim \mathsf{N}(\mathbf{0}, \boldsymbol{\Sigma}^g)$. Note that each unit $i = 1, \ldots, n$ has to be treated separately while using parameters specific to cluster to which unit i currently belongs.

A unique treatment is required for $\mathbf{c}_r^g$ parameter of ordered intercepts for ordinal outcomes. Derivation of Newton-Raphson method leads to an algorithm that does not necessarily guarantee the ordinality of elements of $\mathbf{c}_r^g$. Hence, we propose the following transformation

$$
\begin{aligned}
c_1 &= a_1 & a_1 &= c_1, \\
c_2 &= a_1 + e^{a_2} & a_2 &= \log(c_2 - c_1), \\
c_3 &= a_1 + e^{a_2} + e^{a_3} & a_3 &= \log(c_3 - c_2), \\
\vdots &= \vdots & \vdots &= \vdots \\
c_{K-1} &= a_1 + \sum_{k=2}^{K-1} e^{a_k} & a_{K-1} &= \log(c_{K-1} - c_{K-2}).
\end{aligned}
$$

There is one to one correspondence between the ordered intercepts $\mathbf{c}_r^g$ and newly defined parameter $\mathbf{a}_r^g$, elements of which are no longer restricted. Therefore, we set multivariate normal prior over parameter $\mathbf{a}_r^g$ and apply Metropolis step to it with the use of proposal distribution found by Newton-Raphson method (the transformation preserves smoothness), which now does not face any limitations. After sampling $\mathbf{a}_r^g$ we transform it back to obtain ordered intercepts $\mathbf{c}_r^g$.

In order to fully approximate posterior distribution of parametric functions of allocation probabilities $p_{i,g}(\boldsymbol{\theta})$ we would need to efficiently compute integral (1). Nevertheless, if we settle for posterior mean only, we can utilize full-conditional (also given $\mathbf{B}_i$) classification probabilities that are already calculated for each sampled state in order to obtain new U_i indicators. By the fact that

$$
\int \mathsf{P}\left[U_i = g | \mathbf{Y}_i; C_i, \boldsymbol{\theta}\right] \cdot p(\boldsymbol{\theta} | \mathbf{Y}_i; C_i)\, \mathrm{d}\boldsymbol{\theta} = \int \int \mathsf{P}\left[U_i = g | \mathbf{Y}_i; C_i, \boldsymbol{\theta}, \mathbf{B}_i\right] \cdot p(\boldsymbol{\theta}, \mathbf{B}_i | \mathbf{Y}_i; C_i)\, \mathrm{d}\mathbf{B}_i \mathrm{d}\boldsymbol{\theta},
$$

we can approximate the posterior mean of $p_{i,g}(\boldsymbol{\theta})$ by arithmetic mean of full-conditional probabilities. This approximation can then be used for classifying units into cluster with highest posterior mean probability. However, without full exploration of posterior distribution we can hardly create more sophisticated classification rules that would adequately account for possible indecisiveness.

5 Application to EU-SILC

The Czech subset of EU-SILC dataset gathered between 2005 and 2018 consists of $n = 23\,360$ households that were followed for exactly $n_i = 4$ consecutive years which is induced by rotational design. Each year a quarter of the followed households is dropped to be replaced by a set of new households. For the analysis we primarily use data gathered on household level, however, on special occasions we use gathered personal data to create a new indicator that summarizes the whole household. Outcomes of interest are listed below with respect to their type:

- Numeric outcomes (used on log-scales)
  - `HX090` - Equivalised total disposable income [€/year]
  - `HS130` - Lowest monthly income to make ends meet (to pay for its usual necessary expenses) [€/month]

- Binary outcomes (Yes / No)
  - `HS040` - Capacity to afford paying for one week annual holiday away from home
  - `HS060` - Capacity to face unexpected financial expenses
- Ordinal outcomes (the higher, the less problematic)
  - `HS120` - Ability to make ends meet (self-evaluation by respondent)
    1. with great difficulty
    2. with difficulty
    3. with some difficulty
    4. fairly easily
    5. easily
    6. very easily
  - `HS140` - Financial burden of the total housing cost (self-evaluation by respondent)
    1. a heavy burden
    2. a slight burden
    3. not a burden at all
- Categorical outcomes (Yes / No - cannot afford / No - other reason)
  - `HS090` - Do you have a computer?
  - `HS110` - Do you have a car?

The dataset also offers plenty of potential regressors to be used in the fixed part of the model.

- *Time* is the most important one due to the economical crisis that has struck during the follow-up time. This had a great impact on the households in several different ways. Hence, we decided for quadratic spline parametrization with one inner knot to allow for the change in evolution.
- *Equivalised household size* expresses how large the household is while taking age into consideration. The head of the household (respondent) has unit weight, while other members have either 0.5 or 0.3 depending on their age - older or younger than 14, respectively. We keep it in a linear parametrization as one of the fixed effects.
- *Level of urbanisation* was originally divided by the population density and minimum population into three categories: densely populated area, intermediate area and thinly-populated area. However, the capital city Prague behaves in many aspects very distinctly. For that reason we created fourth category dedicated to households in Prague exclusively.
- *The highest education level achieved* within the whole household rarely attains the lowest possible option of primary education. The most common one - secondary education was divided into lower, upper and post secondary categories. The highest group - tertiary education - are households where at least one member has some kind of university degree.
- *Presence of student* or *baby* are binary indicators whether some household member currently attends any educational institution or is younger than 3 years, respectively.

Usage all of the above suggested regressors creates the fixed part of the predictor dependent on 13-dimensional vector of $\boldsymbol{\beta}$ coefficients plus an intercept term for non-ordinal outcome. Each of the outcomes shares the same structure of the fixed effects but is supposed to have his own group-specific $\boldsymbol{\beta}_r^g$. On the other hand, we try to keep random structure as simple as possible to not to increase the dimension of $\boldsymbol{\Sigma}$ (common to all clusters). Therefore, a simple random intercept term for each considered outcome suffices for our purposes of joint modelling. Then the dimension of $\mathbf{B}_i$ is the lowest possible and corresponds to the number of considered outcomes, which means 8 if all outcomes suggested above are used.

During the analysis we treat the number of groups G to be fixed and known in advance. However, nothing stops us from trying several potential values of $G \in \{2,3,4,5,6\}$ and choose the one that fits best according to cluster interpretations. In [6] we also used posterior distribution of deviance $D(\boldsymbol{\theta}; \mathbf{Y}_1, \dots, \mathbf{Y}_n) = -2 \sum_{i=1}^{n} \log f(\mathbf{Y}_i; C_i, \boldsymbol{\theta})$. However, this approach is as computationally demanding as full exploration of posterior distribution of all classification probabilities $p_{i,g}(\boldsymbol{\theta})$, which for Markov chain of length M includes $M \cdot n \cdot G$ approximations of integral (1).

We classify households into the cluster with the highest posterior mean of classification probabilities (2) only when this estimate dominates other probabilities by a certain margin, e.g. 0.1. Apart from cluster characteristics given by specific fixed effects $\boldsymbol{\beta}_r^g$ we can relate the clustering with other available data. A very interesting comparisons are with the type of the household (its family composition) and with the poverty indicator (whether a household has equivalised total disposable income below 60 % of the median). It shows that our clustering method provides an alternative and more elaborated approach for identification of households endangered by poverty.

6 Conclusion

Although, replacement of latent variable approach by GLMM brings several difficulties with respect to the sampling process for estimation, we were able to circumvent them by the use of Metropolis proposal steps combined with Newton-Raphson method. For this effort we were rewarded by significantly improved convergence properties. However, some of the difficulties still remain in the integration (1). For example, when we want to explore posterior distribution of classification probabilities or deviance in more depth. Although, suggested solution by multivariate version of adaptive Gauss-Hermite quadrature using Laplace approximation (Breslow and Clayton, [1]) can handle the integration quite satisfactorily, the enormous amount of its performance significantly prolongs the computational time.

The next step in the improvement could be to replace generality of covariance matrix $\mathbf{\Sigma}$ by some block structure to make the dependence clearer and to allow for much higher dimensions of random effects and possibly much larger set of outcomes of interest. The GLMM framework and the use of Newton-Raphson method also opens door for other types of models for numeric outcomes, choice of different link function than logit or even for addition of count-type outcomes.

In terms of real data application, our methodology can identify households of both regular and extraordinary behaviour. Classical poverty indicators tend to be one-dimensional and do not take all household aspects into consideration. Our method can absorb diverse kinds of information and then provide a more complex insight. One of the advantages of being still a fully parametric model is that we can describe exactly how the clusters differ among themselves.

Acknowledgements

This research was supported by the Czech Science Foundation (GAČR) grant 19-00015S.

References

[1] Breslow, N. E. and Clayton, D. G. (1993). Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association*, **88**, 9 – 25.

[2] Brooks, S., Gelman, A., Jones, G. and Meng, X. (2011). *Handbook for Markov chain Monte Carlo*. Taylor & Francis.

[3] Fraley, C. and Raftery, A. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, **97**, 611 – 631.

[4] Jiang, J. (2007). *Linear and Generalized Linear Mixed Models and Their Applications*. Springer-Verlag New York.

[5] Laird, N. M. and Ware, J. H. (1982). Random-Effects Models for Longitudinal Data. *Biometrics*, **38**, 963 – 974.

[6] Vávra, J. and Komárek, A. (2020). Identification of Temporal Patterns in Income and Living Conditions of Czech Households: Clustering Based on Mixed Type Panel Data from the EU-SILC Database. *Conference Proceedings of 38th International Conference on Mathematical Methods in Economic (MME2020)*, 612 – 617.