MOX-Report No. 41/2024

Ordinal Mixed-Effects Random Forest

Bergonzoli, G.; Rossi, L.; Masci, C.

MOX, Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox-dmat@polimi.it https://mox.polimi.it

Ordinal Mixed-Effects Random Forest

Giulia Bergonzoli *[1], Lidia Rossi †[1,2], and Chiara Masci ‡[1]

[1]MOX Laboratory, Department of Mathematics, Politecnico di Milano, Milano, Italy

[2]Department of Management, Economics and Industrial Engineering, Politecnico di Milano, Milano, Italy

Abstract

We propose an innovative statistical method, called Ordinal Mixed-Effect Random Forest (OMERF), that extends the use of random forest to the analysis of hierarchical data and ordinal responses. The model preserves the flexibility and ability of modeling complex patterns of both categorical and continuous variables, typical of tree-based ensemble methods, and, at the same time, takes into account the structure of hierarchical data, modeling the dependence structure induced by the grouping and allowing statistical inference at all data levels. A simulation study is conducted to validate the performance of the proposed method and to compare it to the one of other state-of-the art models. The application of OMERF is exemplified in a case study focusing on predicting students performances using data from the Programme for International Student Assessment (PISA) 2022. The model identifies discriminating student characteristics and estimates the school-effect.

Keywords: Mixed-effects models, Ordinal models, Tree-based methods, Random forest, Learning analytics.

*Email: giulia.bergonzoli@gmail.com
†Email: lidia.rossi@polimi.it
‡Email: chiara.masci@polimi.it

1 Introduction

Ordinal-scale observations can be categorized into a finite and ordered set of discrete categories. Distances between categories can be uneven or unknown. These scales are typically constructed by condensing continuous variables into a set of distinct categories. An ordinal variable can be considered quantitative since each level of the scale denotes a greater or lesser magnitude of a particular characteristic compared to another level. (Agresti, 2010). The growing importance of ordinal categorical data shows a clear positive trend, driven by the increasing use of surveys and tests (Yang et al., 2020). This type of data is useful to collect detailed information, especially in fields like market research, public opinion analysis, and healthcare, where assessing opinions, preferences, and responses is extremely important. Order becomes relevant when the categories take on meanings related to strength of opinion, agreement (as in a Likert-type response) or frequency. An explanatory example is the case where a response variable takes on four possible values: (1) strongly disagree, (2) disagree, (4) agree, (5) strongly agree. There is a natural order in the response possibilities.

As data collection expands into various areas, there is a bigger need to model ordered data. Ordinal classification, often known as ordinal regression (McCullagh, 1980), represents a type of multi-class classification where there is an inherent ordering relationship between the classes, but where there is not a meaningful numeric difference between them. This paper proposes an innovative tool for ordinal classification that extends the use of random forest (Breiman, 2001) to the case of ordinal responses and hierarchical observations (Pinheiro and Bates, 2006). The proposed method, called Ordinal Mixed-Effects Random Forest (OMERF), fits into the context of tree-based mixed-effects models (Hajjem et al., 2011; Sela and Simonoff, 2012; Hajjem et al., 2014, 2017; Pellagatti et al., 2021). In particular, the algorithm we implement disentangles the estimations of fixed and random effects, by iteratively fitting

(i) a random forest (Breiman, 2001), ignoring the grouped data structure, and (ii) a cumulative mixed-effects model (Grilli and Rampichini, 2011; Tutz and Hennevogl, 1996), based on the residuals of the random forest structure. A final mixed-effects random forest is reported. To the best of our knowledge, this is the first time that a multilevel random forest for ordinal response is constructed.

The paper is structured as follows. Section 2 conducts a review of the literature related to analogous methods. Section 3 articulates the OMERF method, outlining its theoretical foundations and its implementation. Section 4 reports a simulation study in which we test OMERF and compare it to other counterpart methods. Section 5 delves into a real-world case study, in which the efficacy of the proposed method is proved through the application to data from the Programme for International Student Assessment (PISA) 2022. The aim is the model students' mathematical performance levels, considering students' nested structure within schools. Ultimately, Section 6 is dedicated to highlighting conclusions and fostering a discussion.

All the analysis are performed using R software (R Core Team, 2022) and all the R codes for the OMERF algorithm and for both simulation and case study are available in the following Github repository: `https://github.com/giuliabergonzoli/OMERF`.

2 Literature review

The method proposed in this study takes inspiration from tree-based mixed-effects models, which focus on incorporating tree-based methods (Breiman, 2017) and their ensembles (Breiman, 2001) into the framework of mixed-effects models (Pinheiro and Bates, 2006). Tree-based methods are popular for their ability to capture complex and nonlinear relationships. When declined into the mixed-effects models framework, they can handle both nested and longitudinal data. The first are data that present a hierarchical structure, the second refer to the situation where repeated observations are available for each sampled object. Nested data are not independent and identically distributed (*i.i.d.*) as assumed in classical regression and classification models, but their distribution depends on their grouping structure. Analysing and disentangling the effects associated to each level of the hierarchy enables a deeper understanding and investigation of the regression dynamics, thereby increasing the comprehension of the phenomenon described by the data. It is important to account for this structure, as it can provide significant insights that might otherwise be neglected, enabling the quantification of the portion of variability in the response variable that is attributable to each level of grouping.

Tree-based mixed-effects models developed in statistical literature can be categorized into two groups: the first focuses on Gaussian responses, making it unsuitable for classification tasks, while the second group extends its applicability to non-Gaussian responses and is suitable for addressing classification problems. The works in (Sela and Simonoff, 2012; Hajjem et al., 2011, 2014) pertains to the first collection of models. In (Sela and Simonoff, 2012), the Random Effects Expectation-Maximization (RE-EM) tree is implemented; while in (Hajjem et al., 2011), the authors propose the Mixed-Effect Regression Tree (MERT) model. They both are extensions of conventional regression trees to account for clustered and longitudinal data, substituting the fixed effect component of a linear mixed-effects model with a tree structure. Then, trying to improve prediction accuracy, a method known as Mixed-Effects Random Forest (MERF), in which the fixed effect component is modelled by a random forest, is introduced in (Hajjem et al., 2014). With regard to the extensions to other types of responses, in (Hajjem et al., 2017), the authors propose the Generalized Mixed-Effects Regression Tree (GMERT), which basically extends the MERT approach to non-Gaussian responses. An alternative, proposed in (Fontana et al., 2021), is the Generalized Mixed-Effects Tree (GMET), that follows a three-step procedure: first, the random-effects are initialised to zero and the systematic component is estimated through a generalized linear model; then, a regression tree is built using the estimated systematic component as dependent variable and, finally, a mixed-effects model is fitted to estimate the random-effects part, using the estimated tree as offset. In (Fokkema et al., 2018), the authors propose an algorithm known as Generalized Linear Mixed-effects Model tree (GLMM tree), which iteratively refines the estimates of a generalized linear model tree and a mixed-effects model until convergence is achieved. Lastly, in (Speiser et al., 2020), a decision tree method for modeling clustered and longitudinal binary outcomes within a Bayesian framework, called Binary Mixed Model tree (BiMM tree), is introduced. A step forward in the field of tree-based aggregated models has been done in (Pellagatti et al., 2021), in which the authors implement the Generalized Mixed-Effects Random Forest (GMERF), thus extending for the first time random forest, and not only simple trees, to deal with hierarchical data, both for regression and classification (for any response variable in the exponential family).

Contributing to this branch of the literature, the current research work proposes a novel method called Ordinal Mixed-Effects Random Forest (OMERF), that is inspired by the GMERF model, but extends the multilevel random forest approach to deal with ordinal data, namely, with ordinal regression problems.

Concerning the statistical literature about ordinal data, one of the early contributions to classification techniques for ordinal data can be found in (McCullagh, 1980), where a regression model is introduced. Moving to a multilevel setting, the random effects cumulative model, described in (Tutz and Hennevogl, 1996), addresses ordinal regression models as special cases of multivariate generalized linear models, adjusting them to include random effects in the linear predictor. With regard to the implementation of nonlinear ordinal models, some attempts can be found in the literature. The work in (Tutz, 2003) proposes an extension of ordinal regression through the generalization of the additive model by incorporating nonparametric terms; in (Shashua and Levin, 2002) the authors introduce a generalised formulation for the support vector machine for ordinal data. The ordinal random forest method is presented in (Hornung, 2020), which is a random forest-based prediction method for ordinal response variables. Finally, in (Tutz, 2022), an extension with score-free recursive partitioning and ensembles that include parametric models is proposed.

Within this literature, OMERF, combining random effects cumulative models (Tutz and Hennevogl, 1996) with the ordinal random forest (Hornung, 2020), is the first tree-based method for nested data and ordinal responses and paves the way for a new class of models of increasing interest.

3 Model and Methods

In this section, a concise overview of cumulative link models for ordinal data (Section 3.1.1) and random forest (Section 3.1.2) is provided. This will serve to set the notation and to understand the OMERF method, detailed in Section 3.2.

3.1 Background and state of the art models

3.1.1 Cumulative link mixed models

Ordinal observations can be expressed through a random variable, Y_j, which takes on the value c when the j-th ordinal observation assumes the c-th category. The support of Y_j includes the integer values from 1 to C, where $C \geq 2$. Cumulative Link Models (CLM) (Agresti, 2010; Ananth and Kleinbaum, 1997; McCullagh, 1980) deal with observations on an ordinal scale and apply an arbitrary link function to link the cumulative probabilities to a linear predictor. Cumulative Link Mixed Models (CLMM) (Grilli and Rampichini, 2011; Tutz and Hennevogl, 1996) extends CLMs to deal with nested observations, by including normally distributed random effects. These models are suited to handle data with a hierarchical structure. Given $\boldsymbol{y}_i = y_{i1}, \dots, y_{in_i}$ the n_i-dimensional response vector for observations in the i-th group, the configuration of a CLMM with a random intercept and Q random slopes in a two-level hierarchy can be written as:

$$\begin{aligned} \eta_{ijc} = g(\gamma_{ijc}) = \theta_c - (\boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{b}_i), \qquad c = 1, \dots, C-1 \\ \boldsymbol{b}_i \sim \mathcal{N}_Q(\boldsymbol{0}, \boldsymbol{\Sigma_b}) \end{aligned} \tag{1}$$

where $i = 1, \dots, I$ is the level 2 (group) index and $j = 1, \dots, n_i$ (with $\sum_{i=1}^{I} n_i = J$) is the nested level 1 index. $\gamma_{ijc} = \mathbb{P}(y_{ij} \leq c) = \pi_{ij1} + \cdots + \pi_{ijc}$ (with $\sum_{c=1}^{C} \pi_{ijc} = 1$ and $\pi_{ijc} = \mathbb{P}(y_{ij} = c)$) is the cumulative probability up to the c-th category for unit j

in group i. η_{ijc} is the linear predictor, $\boldsymbol{x}_{ij}$ is a P-dimensional vector of predictors, to which corresponds the fixed effects vector $\boldsymbol{\beta}$, associated to the entire population, and $\boldsymbol{z}_{ij}$ is a $(Q+1)$-dimensional vector of predictors, to which correspond group-specific parameters $\boldsymbol{b}_i$. The elements y_{ij} are supposed to be conditionally independent on the random effects $\boldsymbol{b}_i$. Moreover, random effects $\boldsymbol{b}_i$, conditionally on the fixed effects, are assumed to be independent and identically distributed with zero mean and a common group covariance matrix $\boldsymbol{\Sigma_b}$. Finally, g is a monotonic, differentiable link function and θ_c are the strictly ordered thresholds (also known as cut-points or intercepts):

$$-\infty \equiv \theta_0 \leq \theta_1 \leq \cdots \leq \theta_{C-1} \leq \theta_C \equiv \infty.$$

In CLMMs, the ordinal response variable y_{ij} with C categories is assumed to be generated by a latent continuous variable y^*_{ij} with a set of $C-1$ thresholds θ^*_c such that $y_{ij} = c \iff \theta^*_{c-1} \leq y^*_{ij} \leq \theta^*_c$. The latent continuous variable is modelled as:

$$y^*_{ij} = \boldsymbol{x}^T_{ij}\boldsymbol{\beta^*} + \boldsymbol{z}^T_{ij}\boldsymbol{b}^*_i + \epsilon^*_{ij} \tag{2}$$

where ϵ^*_{ij} is a level 1 error with standard deviation σ_{ϵ^*}.

Therefore, the cumulative probabilities are $\gamma_{ijc} = \mathbb{P}(y_{ij} \leq c) = \mathbb{P}(y^*_{ij} \leq \theta^*_c) = \mathbb{P}(\epsilon^*_{ij} \leq \theta^*_c - \boldsymbol{x}^T_{ij}\boldsymbol{\beta^*} - \boldsymbol{z}^T_{ij}\boldsymbol{b}^*_i) = g^{-1}(\theta_c - \boldsymbol{x}^T_{ij}\boldsymbol{\beta} - \boldsymbol{z}^T_{ij}\boldsymbol{b}_i)$. The underlying linear model (2) with thresholds θ^*_c and level 1 error ϵ^*_{ij} having distribution function g^{-1} is equivalent to the cumulative model (1) with link function g.

In the context of CLMMs, model parameters are typically estimated using maximum likelihood methods, employing techniques such as Adaptive Gaussian Quadrature, Gauss-Hermite Quadrature, or Laplace approximation to the likelihood function. A Newton-Raphson algorithm updates the conditional modes of the random effects for the subsequent approximations and finally a nonlinear optimization is performed over the fixed parameter set to get the Maximum Likelihood Estimation.

3.1.2 Random forest

Random forests (RFs) are a powerful machine learning algorithm that combines the predictive power of multiple decision trees to make accurate and robust predictions. In particular, random forest for regression (Breiman, 2001; James et al., 2013) are tree-based ensemble methods formed by growing regression trees such that the tree predictor $f(\boldsymbol{x})$ takes on numerical values, and the random forest predictor is formed by taking the average over K of these trees $f_k(\boldsymbol{x})$, $k = 1, \ldots, K$.

The extension for ordinal responses, namely ordinal forests (OF) (Hornung, 2020), involves the construction of regression forests in which traditional class values are replaced by score values. These score values are optimized to improve out-of-bag (OOB) prediction performance, evaluated against a designated measure known as the performance function. This model is based on the assumption that exists a continuous variable, denoted as y^*, underlying the observed ordinal variable y, known or unknown, indicating the ordinal variable's values. Specifically, the relationship dictates that as the value of y^* increases for an observation, so does the corresponding class of the ordinal response variable. Class widths are the widths of J adjacent intervals and can vary between class and class.

3.2 Ordinal mixed-effects random forest

The proposed statistical method, called Ordinal Mixed-Effects Random Forest (OMERF), extends the use of random forest to the analysis of hierarchical data, for categorical ordinal response variables. It models the fixed effects through a random forest, combining them to the random effects obtained using a CLMM, in order to take into account both possible complex functional forms in the fixed effect component and the nested structure of data. The model can be formulated as follow:

$$\begin{aligned}
\eta_{ijc} = g(\gamma_{ijc}) = \theta_c - (f(\boldsymbol{x}_{ij}) + \boldsymbol{z}_{ij}^T \boldsymbol{b}_i) \\
g(\gamma_{ijc}) = logit(\gamma_{ijc}) = \ln(\frac{\gamma_{ijc}}{\gamma_{ijc} - 1}) \\
\gamma_{ijc} = \mathbb{P}(y_{ij} \leq c) \\
\boldsymbol{b}_i \sim \mathcal{N}_Q(\mathbf{0}, \boldsymbol{\Sigma_b}) \\
j = 1, \dots, n_i \qquad i = 1, \dots, I \qquad c = 1, \dots, C-1
\end{aligned} \tag{3}$$

where $f(\boldsymbol{x}_{ij})$ is the unknown and nonlinear structure estimated through the random forest, γ_{ijc} are cumulative probabilities, $\pi_{ijc} = \mathbb{P}(y_{ij} = c) = \mathbb{P}(y_{ij} \leq c) - \mathbb{P}(y_{ij} \leq c-1) = logit^{-1}(\theta_c - (f(\boldsymbol{x}_{ij}) + \boldsymbol{z}_{ij}^T \boldsymbol{b}_i)) - logit^{-1}(\theta_{c-1} - (f(\boldsymbol{x}_{ij}) + \boldsymbol{z}_{ij}^T \boldsymbol{b}_i))$ is the probability that the j-th observation, within the i-th group, falls in the c-th category. Similarly to CLMM model, OMERF model assumes that the random effects $\boldsymbol{b}_i$ and $\boldsymbol{b}_{i'}$ are independent for $i \neq i'$. The fixed component is described by a RF object and, consequently, its exploration relies on familiar tools such as partial plots and variable importance plots.

The OMERF algorithm is inspired by the one proposed in (Pellagatti et al., 2021) and estimates fixed and random effects by following an iterative procedure in which the two components are estimated separately. To perform this estimation, it can be observed that if the random effects were known in advance, a RF could be fitted to estimate the fixed part $f(\boldsymbol{x}_{ij})$ by using $\eta_{ijc} + \boldsymbol{z}_{ij}^T \times \boldsymbol{b}_i$ as dependent variable. Similarly, if the population-level effects were known, the random effects could be estimated using a CLMM with the response corresponding to η_{ijc} and using $f(\boldsymbol{x}_{ij})$ as an offset of the model. Since neither of them is known, an iterative approach that alternates between estimating the RF for the fixed component and estimating the CLMM for the random one is employed. Convergence is considered achieved when the difference between the random effects estimates in two consecutive iterations is less than a predetermined tolerance.

The pseudo-code outlining this estimation process is provided in Algorithm 1. The random forest model is constructed using the R package *randomForest* (Liaw and Wiener, 2002), which implements the original algorithm described in (Breiman, 2001). Meanwhile, the CLMM is built using the *clmm* function from the R package *ordinal* (Christensen, 2022). The CLMM model allows for different offsets in the formula and scale effects, which are considered as components of the linear predictor that are known in advance and thus they require no parameter to be estimated from the data. The implemented method in particular will make use of an offset modelled as:

$$\begin{gathered}\eta_{ijc} = g(\gamma_{ijc}) = \theta_c - \boldsymbol{z}_{ij}^T \boldsymbol{b}_i - offset_{ij}, \\ j = 1, \ldots, n_i \quad i = 1, \ldots, I \quad c = 1, \ldots, C-1\end{gathered} \tag{4}$$

where $offset_{ij} = f(\boldsymbol{x}_{ij})$, i.e., the random forest estimates of the fixed component.

Addressing the initialization of the unknown systematic component η_{ijc} represents a delicate challenge, as it cannot be directly inferred from the data. To tackle this issue, we utilize a traditional Ordinal Forest model incorporating covariates of the fixed component as predictors and the vector with ordinal categorical responses y_{ij} as target. This model is used to estimate the cumulative probabilities γ_{ijc} and, then, the inverse link function g^{-1} is applied to initialize $\eta_{ijc} = g^{-1}(\gamma_{ijc})$. The implementation of the ordinal forest can be found in the *ordinalForest* R package (Hornung, 2020). This approach represents an enhancement compared to similar methods like GMERF (Pellagatti et al., 2021), which employs a GLM for initializing η_{ij}, thereby missing out on the benefits of non-parametric methods, at the step 0 of the algorithm. Specifically, using a GLM for initialization fails to capture potentially nonlinear trends and interactions, a capability that OMERF aims to achieve.

Once OMERF has been fitted, to make predictions for a new observation $[\boldsymbol{x}_{ij}; \boldsymbol{z}_{ij}]$, the following formula is employed: $\hat{\eta}_{ijc} = \hat{\theta}_c - (\hat{f}(\boldsymbol{x}_{ij}) + \boldsymbol{z}_{ij}^T \hat{b}_i)$. Here, $\hat{f}$ represents the random forest model estimated by the algorithm, $\hat{b}_i$ is the vector of random effects associated to the i-th group, and $\hat{\theta}_c$ is the threshold associated to each predicted category c.

Algorithm 1 OMERF

Input:

$y-$ vector with ordinal categorical responses y_{ij}

$cov-$ data frame with all covariates

$group-$ vector with the grouping variable for each observation

$xnam-$ vector with names of the covariates to be used as fixed effects

$znam-$ vector with names of the covariates to be used as random effects

b_0- optional matrix of initial values for each $\underline{b}_i$

$toll-$ threshold to decide whether our estimation converged or not (default value of 0.05)

$itmax-$ maximum number of iterations (default value of 100)

$Z \leftarrow (1;cov[znam])$: it includes also the random intercept

Initialize b to a matrix of zero (if b_0 is not given): each column $b[,i]$ of b will be the i-th random coefficients $\underline{b}_i$

$all.b[0] = b$

fit a Ordinal Forest model using y as response and cov as matrix of covariates

$eta \leftarrow$ estimated η_{ijc} by the Ordinal Forest model

$it \leftarrow 1$

while $it < itmax$ **and not** $conv$ **do**

    $targ \leftarrow eta + Z \times b$

    fit a random forest model using $targ$ as target and cov as predictor matrix

    $fx \leftarrow$ fitted values of the forest model

    fit the CLMM model $\eta_{ijc} = \theta_c - \underline{z}_{ij}^T \underline{b}_i - offset_{ij}$, with $offset_{ij} = fx$

    $all.b[it] = b \leftarrow$ the estimated b fromm CLMM model

    $M \leftarrow \max(abs(b - all.b[it-1]))$

    $(i,j) \leftarrow \text{argmax}(abs(b - all.b[it-1]))$

    $tr \leftarrow M/all.b[it-1](i,j)$

    if $tr < toll$ **then**

        $conv \leftarrow$ **true**

    else

        $conv \leftarrow$ **false**

    end if

    $it{+}{+}$

end while

if not $conv$ **then**

    give a warning

end if

Output:

$clmm.model-$ the final CLMM model fitted

$forest.model-$ the final forest model fitted

$b-$ the final estimation of the random coefficients

$it-$ the number of iterations

4 Simulation study

In this section, we conduct a simulation study to test OMERF and compare it with similar classification methods under various simulation settings. In Section 4.1, we describe the design of the data generating process (DGP), while, in Section 4.2, results are analysed, in order to highlight strengths and weaknesses of OMERF.

4.1 Simulation design

To sample ordered categorical data, we make use of the function *genOrdCat*() from the R package *simstudy* (Goldfeld and Wujciak-Jens, 2020). This function takes as input an underlying (continuous) latent process w_{ij} as the basis for data generation. Assuming that probabilities are determined by segments of a logistic distribution, it defines the ordinal mechanism using thresholds along the support of the distribution. In case of C possible responses, there will be $C - 1$ thresholds. The area under the logistic density curve of each of the C regions defined by those thresholds represents the probability of each possible response tied to that region.

In our simulation, we set $C = 3$ and we generate the underlying latent process as:

$$w_{ij} = f(\boldsymbol{x}_{ij}) + \sum_{q=0}^{Q} z_{qij}^T b_{qi} \qquad j = 1, \dots, n_i \qquad i = 1, \dots, I \tag{5}$$

where f is the fixed component functional form which takes in input the P-dimensional vector of fixed effects covariates $\boldsymbol{x}_{ij}$ and $\sum_{q=0}^{Q} z_{qij}^T b_{qi}$ is the random component. The ordinal response Y is generated from w_{ij} by using the function *genOrdCat*() assuming balanced categories.

Regarding the fixed effects part, we use a variation of the simulation design proposed in (Pellagatti et al., 2021). The design for f incorporates both a linear component and a tree-like component, along with interactions among the covariates. This approach allows to simulate scenarios with highly diverse structure, which will challenge the flexibility and adaptability of our method.

Specifically, $P = 7$ fixed effects covariates are taken into account and f is modelled as follows:

$$f(X_1, \dots, X_7) = \alpha(3 + 7X_1^2 - 5X_2 + X_2X_3^2) + \beta tree(X_4, X_5, X_6) \tag{6}$$

where α and β represent two design parameters employed to regulate the importance given to the linear and tree-based components in the various DGPs. The function $tree(X_4, X_5, X_6)$ follows the tree like stucture outlined in Figure 1. The variable X_7 is included even though it is not significant, in order to assess whether the algorithm is influenced by it. Indeed, while all of the seven variables are being used as predictors in the compared models, only the first six of them are actually used to generate f.

The seven variables are generated randomly in accordance with the following distributions: $X_1, X_2, X_3 \sim N(0, 1)$; $X_4 \sim U(-3, 3)$; $X_5 \sim U(-6, 6)$; $X_6 \sim U(-5, 5)$; $X_7 \sim U(-4, 4)$.

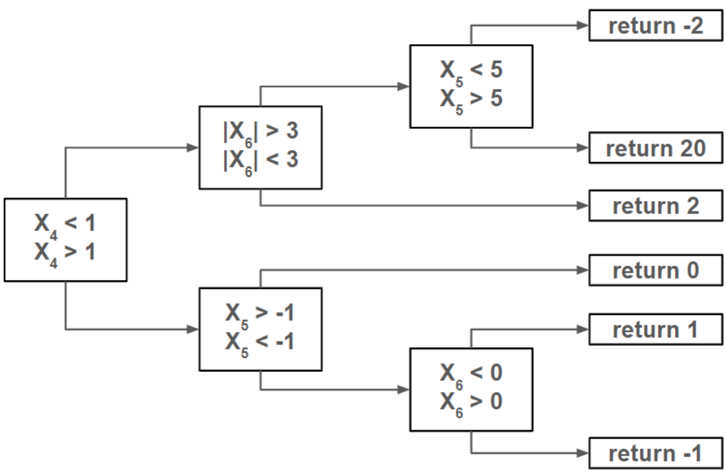


Figure 1: Tree-like structure $tree(X_4, X_5, X_6)$ of the fixed effects part in Equation 6.

The random effects are drawn from a normal distribution for two distinct scenarios:

- *Random intercept only* : $\sum_{q=0}^{Q} z_{qij}^T b_{qi} = b_{0i} \sim \mathcal{N}(0, \sigma_1^2)$, where σ_1^2 is a design parameter which gives the possibility to change the variability of the effects in the different simulations. Indeed, within each scenario, two specifications (low and high) for the parameter σ_1^2 are considered in order to account for different levels of magnitude of the between-group variability;

- *Random intercept and slope* : $\sum_{q=0}^{Q} z_{qij}^T b_{qi} = b_{0i} + x_{1ij}^T b_{1i}$, where X_1 has been previously defined and $\boldsymbol{b}_i \sim \mathcal{N}_2(0, \Sigma)$, with $\Sigma = diag(\sigma_1^2; \sigma_2^2)$. It is important to note that the random effects b_{0i} and b_{1i} are treated as independent for any given value of i, and, as in the previous scenario, σ_1^2 and σ_2^2 regulate the random effects variance. In this setting, the covariate X_1 is not assumed to have a fixed effect that applies uniformly to the entirety of observations. Instead, its association to the response is considered group-specific, meaning that X_1 is assumed to have different effects across observations belonging to distinct groups.

A two-level data structure of $I = 10$ groups with $n_i = 100$ observations each is simulated, for a total number of 1000 units. Note that for simplicity, an equal number of observations for each group is taken to ensure a balanced dataset. However, the model is, of course, capable of handling datasets with varying group sizes.

In addition to this simulation design, a full linear DGP is implemented in order to observe how OMERF performs in a parametric context. In this case, the fixed component includes only the first three variables previously described and is defined as:

$$f(X_1, \ldots, X_3) = 3 + 7X_1 - 5X_2 + X_2X_3. \tag{7}$$

Only the case of a single, normally distributed random intercept is considered: $b_{0i} \sim \mathcal{N}(0, \sigma_1^2)$.

The parameters listed in Table 1 are chosen in order to obtain balanced datasets regarding the classes of the ordinal response. Moreover, several combinations of parameters are tested to observe how the model reacts in the case of a more polynomial or more tree-like systematic component and with the variance related to random effects being higher or lower. Thus, a total of 10 DGPs summarised in Table 1 are obtained. The main objective of the designed DGPs is to show how our algorithm is able to capture the heterogeneity across groups and the structure of the fixed component.

OMERF performance is evaluated in comparison to other three models:

- CLM and CLMM, from the *ordinal* R package (Christensen, 2022), which are expected to perform better in a full linear context, but worse in cases of more complex structures, and CLM is expected not to grasp the hierarchical structure of data;

- Ordinal random forest, from the *ordinalForest* R package (Hornung, 2020), which, as typical of ensemble tree-based models, is able to capture nonlinear relationships, but not to catch the nested structure.

These methods are chosen in order to, at least to some extent, include a range of both cumulative link and tree-based models.

The choice of appropriate performance metrics for ordinal models is not straightforward and is an area of research still not widely explored (de Raadt et al., 2021). Therefore, we compare the performance of the tested methods by employing multiple godness of fit (gof) metrics. We consider the accuracy, that is designed to deal with categorical data and allows to track the percentage of correctly classified observation, but not the error severity, and the Mean Square Error (MSE), that treats the ordinal scale as real number Gaudette and Japkowicz (2009). Furthermore, we incorporate two indexes that assess the similarity between two classifications of the same objects by quantifying the agreement proportions between the two partitions. These are the

Adjusted Rand Index (Hubert and Arabie, 1985) and Cohen's kappa (Cohen, 1960). Both of these indices have a range from -1 to 1. Positive values in this range indicate agreement between the two sets, with 1 denoting perfect agreement. Negative values imply disagreement, and the magnitude of the negative value reflects the extent of this disagreement. A value equal to 0 suggests that the agreement is no different from what would be expected by chance. Lastly, we select among the recent developments in the field the two indexes implemented by J. S. Cardoso (Cardoso and Sousa, 2011) and E. Ballante (Ballante et al., 2022). These last two indexes are novel metrics specifically adapted to ordinal data classification problems, they allow values in the range [0;1] with the optimal value in 0.

	$f(\boldsymbol{x}_{ij})$	α	β	σ_1^2	σ_2^2
DGP 1	$\alpha(3+7X_1^2-5X_2+X_2X_3^2)+\beta tree(X_4,X_5,X_6)$	0.3	0.7	1	-
DGP 2	$\alpha(3+7X_1^2-5X_2+X_2X_3^2)+\beta tree(X_4,X_5,X_6)$	0.7	0.3	1	-
DGP 3	$\alpha(3+7X_1^2-5X_2+X_2X_3^2)+\beta tree(X_4,X_5,X_6)$	0.3	0.7	5	-
DGP 4	$\alpha(3+7X_1^2-5X_2+X_2X_3^2)+\beta tree(X_4,X_5,X_6)$	0.7	0.3	5	-
DGP 5	$\alpha(3+7X_1^2-5X_2+X_2X_3^2)+\beta tree(X_4,X_5,X_6)$	0.3	0.7	0.3	0.5
DGP 6	$\alpha(3+7X_1^2-5X_2+X_2X_3^2)+\beta tree(X_4,X_5,X_6)$	0.7	0.3	0.3	0.5
DGP 7	$\alpha(3+7X_1^2-5X_2+X_2X_3^2)+\beta tree(X_4,X_5,X_6)$	0.3	0.7	1	1
DGP 8	$\alpha(3+7X_1^2-5X_2+X_2X_3^2)+\beta tree(X_4,X_5,X_6)$	0.7	0.3	1	1
DGP 9	$3+7X_1-5X_2+X_2X_3$	-	-	1	-
DGP 10	$3+7X_1-5X_2+X_2X_3$	-	-	5	-

Table 1: Simulation parameters of both fixed and random effects parts for 10 different DGPs.

4.2 Simulation results

For each of the ten DGPs described in Table 1, we simulate 100 datasets. In order to evaluate the predictive performances of the four compared models, each dataset is randomly split into training and test sets, with a ratio of 80% for training and 20% for testing. Simulation results, in terms of mean and variance of gof metrics computed across the 100 runs, are reported in Table 2[1].

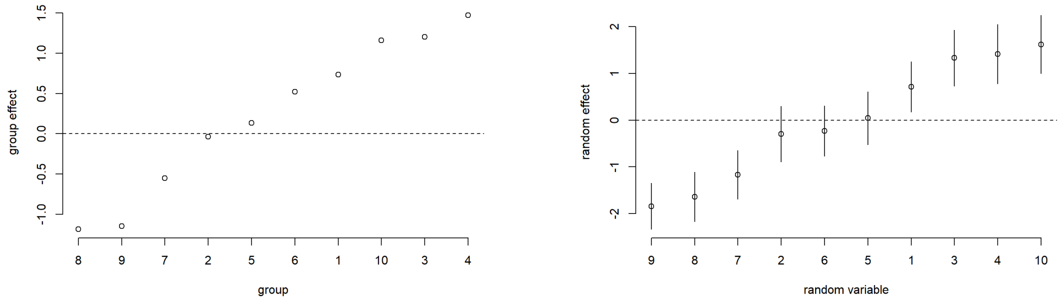
For DGPs from 1 to 4 (i.e., fixed component incorporating both polynomial and tree-based structures and random intercept with varying effect), OMERF consistently emerges as the optimal choice across all reported metrics. Notably, the efficacy gap between tree-based methods (ordinal forest and OMERF) and linear models (CLM and CLMM) can be appreciated both in scenarios with a preponderant tree-like structure and in the ones with a preponderant polynomial structure. The performances of the OMERF method in these scenarios highlights its ability to best capture complex relationships between target and covariates.

With the introduction of a random slope (namely in the DGs 5-8), the models performance remains consistent with those analysed so far. Notably, random forest consistently outperforms linear models, although OMERF does not consistently emerge as the best performer across all indices. Moreover, in all the simulations there appear to be no major differences between the cases with small or large variability of random effects.

On the contrary, in DGPs 9 and 10, in which the predictor is linear, CLM and CLMM tend to perform better with respect to the other models. This result confirms that tree-based methods better capture nonlinear dependencies, but when the data structure is linear parametric, linear models are preferable, additionally yielding results that are usually more easily interpretable.

Overall, results confirm that, in a nonlinear setting, OMERF performs better

[1]OMERF and the ordinal forest are run with default inputs.



(a) Distribution of random intercepts sampled from $\mathcal{N}(0, 1)$.

(b) Distribution of random intercepts with their 95% confidence intervals estimated by OMERF.

Figure 2: Sampled and estimated random intercepts in one of the runs of DGP 1, described in Table 1.

with respect to CLM and CLMM, and slightly better or comparably to ordinal random forest, still having the advantage of extracting knowledge from the nested data structure.

For what concerns the estimation of the predictor, we provide an example of the OMERF output by reporting, in Figures 2 and 3, the results, in terms of fixed and random effects, in one of the runs of DGP 1[2]. Figure 2 reports the sampled (Figure 2(a)) and estimated (Figure 2(b)) random intercepts in one of the runs of DGP 1 and shows how OMERF succesfully manages to capture the heterogeneoty at the group level.

Regarding the fixed effects, Figure 3 shows the net association between the covariates and the response, by means of the partial plots extracted from the RF, giving an insight into the underlying latent process behind the ordinal model. It can be observed that the algorithm captures the quadratic and inverse linear trend in the variables x_1 and x_2, respectively.

[2]The choice of reporting results for a single run is forced by the impossibility of summarizing this type of outcome across the runs.

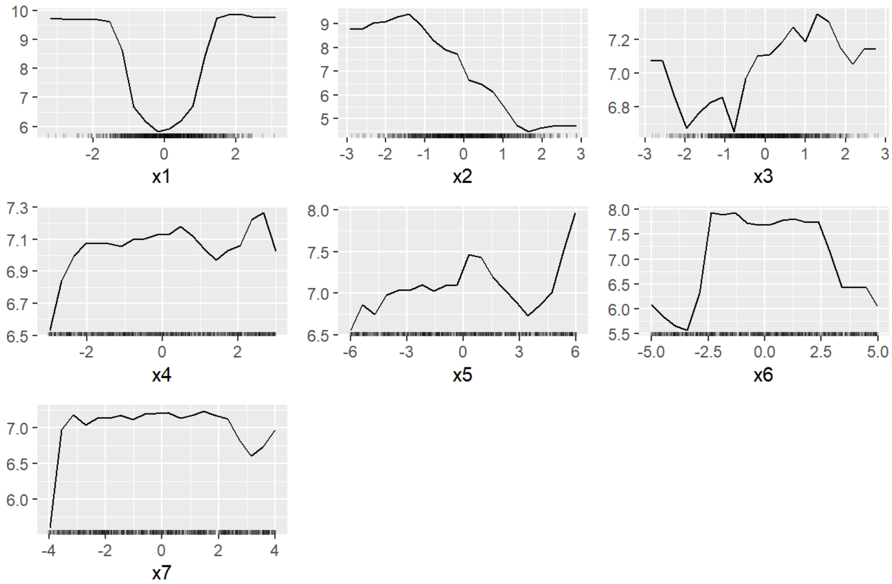


Figure 3: Partial plots for the fixed component of OMERF, for the seven covariates of DGP 1, described in Table 1. The y-axis reports the increment/decrement of the target variable of the random forest in the iterative procedure, given the covariate on the x-axis.

DGP	**Model**	**Acc**	**MSE**	**ARI**	**Cohen's k**	**Cardoso**	**Ballante**
1	clm	0.5729	0.7074	0.1228	0.1650	0.5498	0.2731
		(0.0019)	(0.0109)	(0.0018)	(0.0032)	(0.0022)	(0.0013)
1	clmm	0.5713	0.7127	0.1176	0.1535	0.5503	0.2742
		(0.0018)	(0.0103)	(0.0017)	(0.0031)	(0.0021)	(0.0014)
1	ordforest	0.6491	0.5076	0.2588	0.3346	0.4620	0.1991
		(0.0009)	(0.0046)	(0.0024)	(0.0039)	(0.0013)	(0.0007)
1	omerf	0.6559	0.4506	0.2977	0.3963	0.4619	0.1983
		(0.0011)	(0.0037)	(0.0023)	(0.0028)	(0.0014)	(0.0006)
2	clm	0.7496	0.5047	0.1923	0.2042	0.3811	0.1704
		(0.0008)	(0.0067)	(0.0035)	(0.0036)	(0.0016)	(0.0032)
2	clmm	0.7512	0.5061	0.1812	0.1962	0.3791	0.1806
		(0.0008)	(0.0063)	(0.0033)	(0.0035)	(0.0015)	(0.0040)

2	ordforest	0.8025	0.3148	0.3997	0.3954	0.2909	0.0843
		(0.0006)	(0.0026)	(0.0042)	(0.0047)	(0.0011)	(0.0003)
2	omerf	0.8089	0.2565	0.5330	0.5066	0.2836	0.0899
		(0.0007)	(0.0021)	(0.0036)	(0.0037)	(0.0013)	(0.0004)
3	clm	0.6238	0.6589	0.2182	0.2943	0.5082	0.2314
		(0.0031)	(0.0197)	(0.0043)	(0.0054)	(0.0041)	(0.0025)
3	clmm	0.6236	0.6646	0.2156	0.2893	0.5082	0.2330
		(0.0031)	(0.0202)	(0.0044)	(0.0059)	(0.0041)	(0.0025)
3	ordforest	0.6705	0.5371	0.2853	0.3576	0.4490	0.1829
		(0.0016)	(0.0079)	(0.0038)	(0.0082)	(0.0022)	(0.0010)
3	omerf	0.6748	0.4589	0.3334	0.4237	0.4464	0.1792
		(0.0016)	(0.0050)	(0.0032)	(0.0036)	(0.0020)	(0.0008)
4	clm	0.7532	0.5372	0.2389	0.2725	0.3843	0.2100
		(0.0015)	(0.0143)	(0.0036)	(0.0043)	(0.0031)	(0.0069)
4	clmm	0.7536	0.5377	0.2327	0.2659	0.3839	0.2118
		(0.0015)	(0.0138)	(0.0036)	(0.0047)	(0.0030)	(0.0071)
4	ordforest	0.7953	0.3708	0.3583	0.3722	0.3102	0.0998
		(0.0009)	(0.0059)	(0.0075)	(0.0080)	(0.0019)	(0.0010)
4	omerf	0.8069	0.2865	0.5126	0.5063	0.2914	0.0869
		(0.0011)	(0.0039)	(0.0043)	(0.0036)	(0.0022)	(0.0005)
5	clm	0.5598	0.7261	0.0970	0.1288	0.5604	0.2819
		(0.0009)	(0.0045)	(0.0012)	(0.0026)	(0.0008)	(0.0006)
5	clmm	0.5586	0.7315	0.0940	0.1206	0.5611	0.2851
		(0.0009)	(0.0048)	(0.0014)	(0.0029)	(0.0021)	(0.0009)
5	ordforest	0.6409	0.5002	0.2541	0.3283	0.4659	0.2038
		(0.0009)	(0.0033)	(0.0025)	(0.0033)	(0.0013)	(0.0005)

5	omerf	0.6342	0.4938	0.2645	0.3623	0.4866	0.2166
		(0.0012)	(0.0041)	(0.0033)	(0.0032)	(0.0014)	(0.0005)
6	clm	0.7489	0.4900	0.1803	0.1854	0.3786	0.1653
		(0.0005)	(0.0036)	(0.0038)	(0.0039)	(0.0008)	(0.0025)
6	clmm	0.7503	0.4927	0.1699	0.1812	0.3776	0.1770
		(0.0005)	(0.0041)	(0.0038)	(0.0038)	(0.0008)	(0.0034)
6	ordforest	0.8063	0.2986	0.4187	0.4089	0.2834	0.0824
		(0.0004)	(0.0023)	(0.0040)	(0.0049)	(0.0008)	(0.0003)
6	omerf	0.8104	0.2569	0.5395	0.5119	0.2830	0.0910
		(0.0007)	(0.0021)	(0.0042)	(0.0043)	(0.0013)	(0.0003)
7	clm	0.5696	0.7339	0.1205	0.1691	0.5571	0.2791
		(0.0015)	(0.0133)	(0.0020)	(0.0039)	(0.0019)	(0.0020)
7	clmm	0.5643	0.7491	0.1119	0.1571	0.5630	0.2848
		(0.0018)	(0.0127)	(0.0021)	(0.0045)	(0.0021)	(0.0022)
7	ordforest	0.6414	0.5334	0.2554	0.3254	0.4719	0.2074
		(0.0011)	(0.0047)	(0.0022)	(0.0031)	(0.0015)	(0.0008)
7	omerf	0.6339	0.5121	0.2714	0.3589	0.4882	0.2169
		(0.0016)	(0.0071)	(0.0035)	(0.0036)	(0.0021)	(0.0012)
8	clm	0.7507	0.4943	0.2032	0.2139	0.3774	0.1729
		(0.0008)	(0.0073)	(0.0040)	(0.0040)	(0.0017)	(0.0036)
8	clmm	0.4985	0.4927	0.1908	0.2029	0.3789	0.1809
		(0.0008)	(0.0069)	(0.0042)	(0.0047)	(0.0016)	(0.0040)
8	ordforest	0.8011	0.3235	0.4025	0.3995	0.2947	0.0885
		(0.0005)	(0.0032)	(0.0047)	(0.0041)	(0.0011)	(0.0003)
8	omerf	0.8018	0.2766	0.5195	0.4942	0.2955	0.0965
		(0.0007)	(0.0034)	(0.0032)	(0.0031)	(0.0016)	(0.0004)

9	clm	0.8711	0.1610	0.7291	0.7761	0.1979	0.0419
		(0.0005)	(0.0016)	(0.0021)	(0.0015)	(0.0012)	(0.0001)
9	clmm	0.8716	0.1609	0.7297	0.7769	0.1977	0.0417
		(0.0004)	(0.0014)	(0.0019)	(0.0012)	(0.0009)	(0.0001)
9	ordforest	0.8544	0.2145	0.6858	0.7414	0.2262	0.0484
		(0.0004)	(0.0024)	(0.0022)	(0.0013)	(0.0012)	(0.0002)
9	omerf	0.8389	0.2704	0.7127	0.6167	0.2510	0.0818
		(0.0004)	(0.0028)	(0.0024)	(0.0011)	(0.0011)	(0.0002)
10	clm	0.8756	0.1559	0.7376	0.7824	0.1919	0.0386
		(0.0004)	(0.0012)	(0.0019)	(0.0013)	(0.0009)	(9.1742e-05)
10	clmm	0.8755	0.1559	0.7376	0.7821	0.1920	0.0385
		(0.0004)	(0.0011)	(0.0017)	(0.0012)	(0.0009)	(9.0919e-05)
10	ordforest	0.8420	0.2610	0.6492	0.7154	0.2486	0.0523
		(0.0005)	(0.0031)	(0.0025)	(0.0013)	(0.0012)	(0.0001)
10	omerf	0.8214	0.3340	0.6004	0.6779	0.2845	0.0911
		(0.0005)	(0.0054)	(0.0037)	(0.0015)	(0.0015)	(0.0005)

Table 2: Mean and variances of prediction performances, measured by six indices, of the four compared methods across 100 runs of the 10 DGPs listed in Table 1.

5 Case study

In this section, we delve into a real-world application of the OMERF method in learning analytics.

5.1 The dataset

The data employed in the study concern 15-year-old students attending Italian schools who completed the PISA 2022 survey questionnaire. The dataset comprises information about 10,552 students across 340 schools. After filtering for complete cases and schools with more than 10 students, the processed dataset consists of 7,639 observations, representing students enrolled in 293 schools. For model training and evaluation, a random sample comprising 80% of the observations is designated as the training set, while the remaining 20% constitutes the test set.

The objective of this case study is to evaluate the performance of OMERF in predicting and modelling students' mathematical performance, accounting for student characteristics and attended schools. In accordance with the threshold established by the OECD, the output variable, that is the student PISA score, is categorized into three ordinal levels: the lowest level encompasses students classified in levels 1 or 2, while the highest level includes those achieving levels 5 or 6, with the remaining students falling within the intermediate class [3]. The student-level variables, extracted from the OECD-PISA database, used to predict the mathematics test scores include demographic factors, educational indicators, family background information, factors related to home and school environment, and self-perception attributes. A detailed description of these variables, along with corresponding descriptive statistics, is provided in Table 3. All indicators variables are build by PISA by combining multiple

[3] Levels 5 and 6 are designated to students with high abilities, levels 3 and 4 to students with moderate abilities, and the remaining two levels to those with basic or no abilities. For more details about the PISA proficiency levels, please refer to https://www.oecd.org/pisa/.

responses to questionnaires in numerical indicators.

Variable name	**Variable description**	**Variable type**	**Distribution**
mate3	Output variable, mathematics test score	ordered factor	1:48% 2:44% 3:8%
gender	Student gender (0 = male, 1 = female)	factor	1:53% 0:47%
immig	Student immigration status (0 = native Italian; 1 = 1^{st}-generation immigrant; 2 = 2^{nd}-generation immigrant)	factor	0:89% 1:8% 2:3%
grade	School grade attended (10 = regular student; 9 = late enrolled student; 11 = early enrolled student)	factor	9:10% 10:85% 11:5%
video_games	Indicator of frequency of use at home of video or on-line games	numeric	Mean:3.27; SD:1.59; Range: [1.00;6.00]
internet_quality	Indicator of quality of access to ICT at school	numeric	Mean:-0.18; SD:0.83; Range: [-2.80;2.89]
internet_availability	Indicator of ICT availability outside of school	numeric	Mean:5.70; SD:0.91; Range: [0.00;6.00]
SCHRISK	Indicator of perceived school safety risks	numeric	Mean:0.01; SD:0.82; Range: [-0.46;3.05]
BULLIED	Indicator of being bullied	numeric	Mean:-0.47; SD:0.87; Range: [-1.23;4.69]
BELONG	Indicator of sense of belonging to school	numeric	Mean:0.01; SD:0.89; Range: [-3.26;2.78]
COOPAGR	Indicator of cooperation (agreement)	numeric	Mean:0.12; SD:1.00; Range: [-5.24;6.13]

Continued on next page

Variable name	Variable description	Variable type	Distribution
TEACHSUP	Indicator of perceived teachers support	numeric	Mean:-0.20; SD:1.11; Range: [-2.91;1.56]
FAMSUP	Indicator of perceived family support	numeric	Mean:-0.02; SD:0.93; Range: [-3.01;1.96]
PERSEVAGR	Indicator of perseverance (agreement)	numeric	Mean:0.07; SD:0.98; Range: [-5.91;4.89]
ASSERAGR	Indicator of assertiveness (agreement)	numeric	Mean:-0.03; SD:1.01; Range: [-8.23;7.23]
EMPATAGR	Indicator of empathy (agreement)	numeric	Mean:0.01; SD:0.99; Range: [-6.46;4.69]
EMOCOAGR	Indicator of emotional control (agreement)	numeric	Mean:-0.09; SD:0.98; Range: [-5.17;5.58]
STRESAGR	Indicator of stress resistance (agreement)	numeric	Mean:-0.18; SD:1.00; Range: [-5.26;5.49]
CURIOAGR	Indicator of curiosity (agreement)	numeric	Mean:0.09; SD:0.96; Range: [-4.95;4.18]
study_time	Total time for all homework in all subjects per week	numeric	Mean:3.50; SD:1.51; Range: [1.00;6.00]
HISCED	Highest level of education of parents	numeric	Mean:7.03; SD:2.14; Range: [1.00;10.00]
ESCS	socio-economic family index	numeric	Mean:-0.01; SD:0.87; Range: [-3.23;2.78]

Table 3: Student-level variables extracted from the OECD-PISA database.

5.2 Model results

We run OMERF to predict the student mathematics level `mate3`, that is the ordinal target variable, by including all other variables listed in Table 3 as fixed-effects covariates and considering a random intercept ($Q = 0$) to estimate the school-effect

(Raudenbush and Willms, 1995). We compare the performance of OMERF with the ones of CLM, CLMM and Ordinal Forest[4]. For CLM and Ordinal Forest, the random intercept is not considered.

In Table 4, the predictive performance (computed on the test set) of the four methods are reported.

Model	Acc	MSE	ARI	Cohen's k	Cardoso	Ballante
clm	0.5938	0.4573	0.0739	0.2484	0.5073	0.2835
clmm	0.6699	0.3339	0.1884	0.3963	0.4228	0.2121
ordforest	0.6010	0.4383	0.0914	0.2597	0.4937	0.2607
omerf	0.6444	0.3734	0.1496	0.3546	0.4527	0.2402

Table 4: Prediction performances (computed on the test set) of the four compared methods applied to the real-world case study.

CLMM consistently outperforms other methods across all metrics considered, closely followed by OMERF, which achieves slightly lower but comparable results. This underscores the importance of accounting for the hierarchical structure of the data. The fact that CLMM performs slightly better than OMERF could be explained by the presence of a strong linear relationship between predictors and response. To check that, Figures 4 and 5 illustrate the variable importance plot (VIMP) and the partial plots produced by OMERF, that enable examination of the relationship between predictors and the output variable. From the VIMP (4), we observe that ESCS is the most important predictor, consistent with existing literature. In the partial plot (5), ESCS demonstrates a quasi-linear relationship with the response, potentially explaining the comparable performance of OMERF and CLMM, as linear models can adequately capture such relationships. On the other hand, Figure 5 highlights the nonlinear association between some of the most important variables, such as *EMO-*

[4]OMERF and the ordinal forest are run with default inputs.

COAGR, *CURIOAGR* and *FAMSUP*, and the output variable. For instance, for what concerns the student emotional control (*EMOCOAGR*), we observe a steep increment of the response when this covariate exceeds its average value of 0, while we observe slight variations in the response for both high and low values of the covariate, with high values being associated to higher values of the response. Concerning student curiosity (*CURIOAGR*), we observe a clear increase in the response when the value of this covariate moves from 0 to the right limit of its range, while we observe no significant variations across its negative values. One of the strengths of OMERF is its ability to capture these types of relationships, which CLMM can not. However, the variable importance plot indicates that these covariates are not the most influential variables. This, combined with the quasi-linear relationship between ESCS and the output variable, likely explains the higher performance of CLMM compared to OMERF.

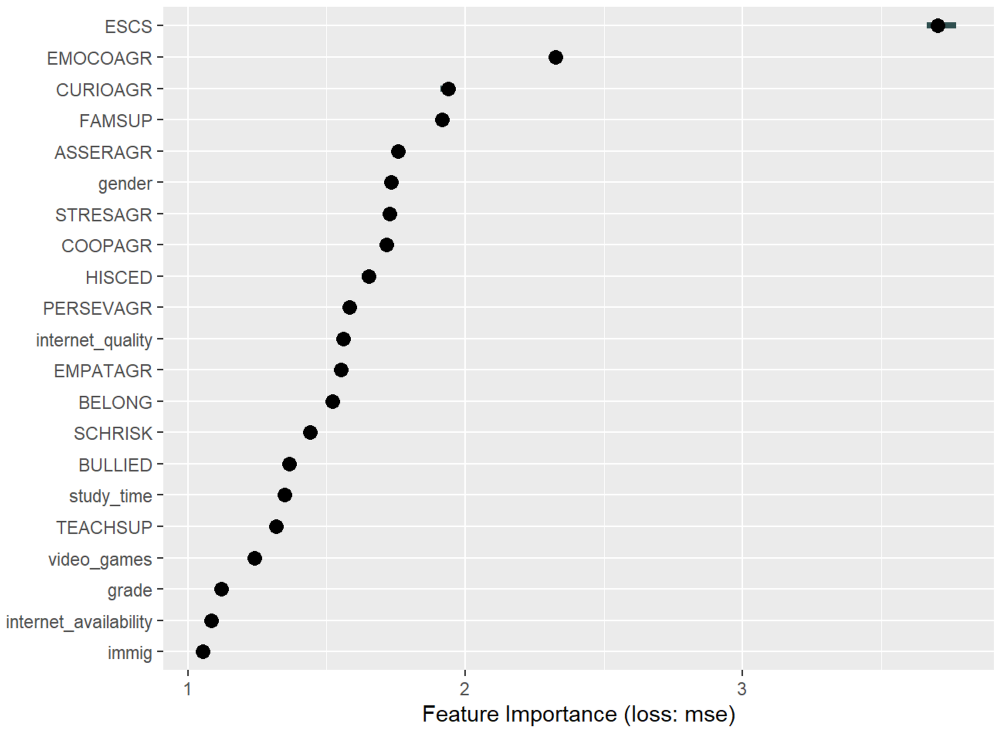


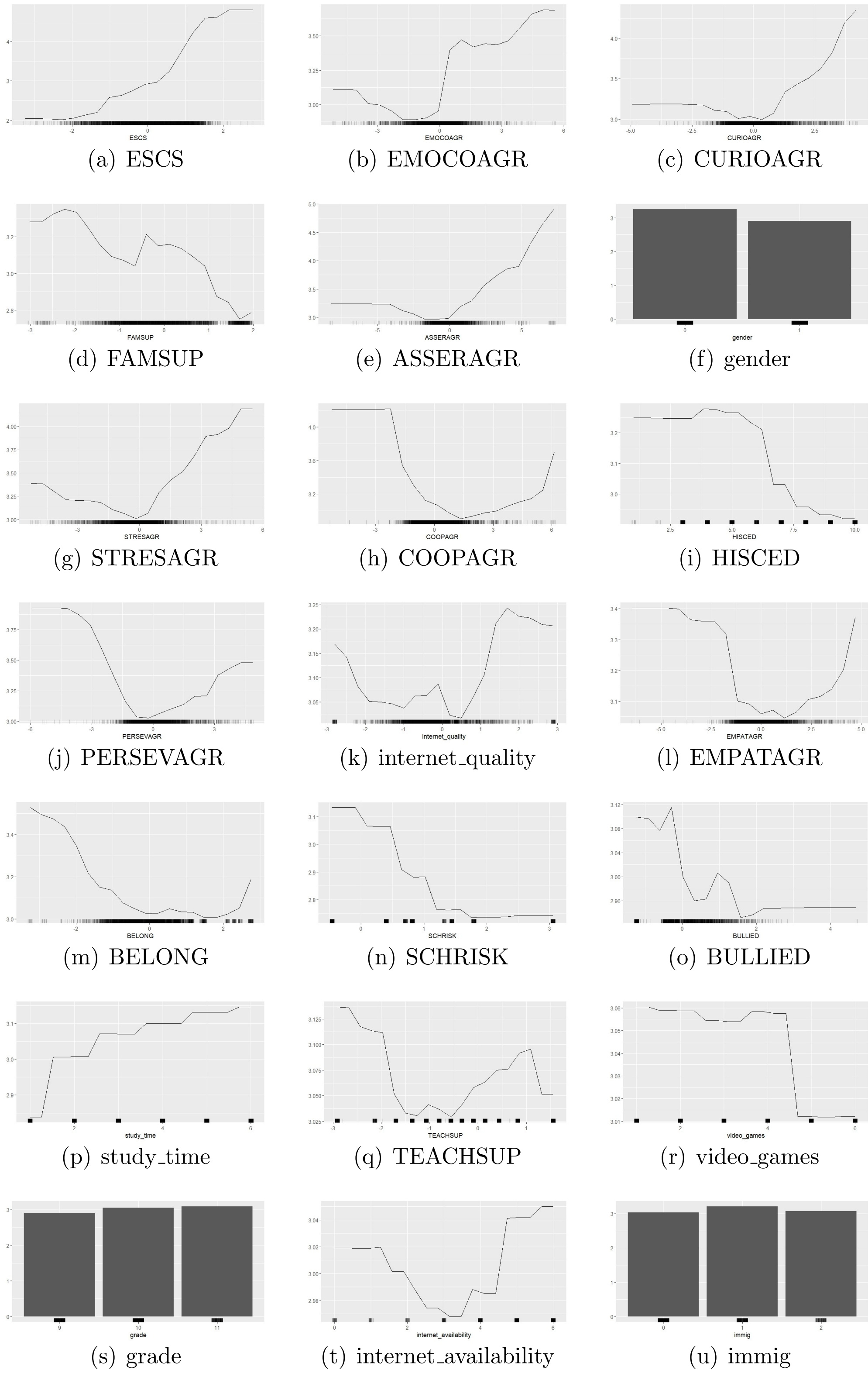
Figure 4: Variable Importance plot for the fixed component of OMERF, in the real-world case study.

The estimated variance of the random effects results to be $\hat{\sigma}^2 = 1.695$ for OMERF, leading to the following Intraclass Correlation Coefficients (ICC) (Grilli and Rampichini, 2011) for the underlying linear model:

$$ICC = \frac{\sigma^2}{\sigma^2 + \pi^2/3} = 0.340.$$

This value of ICC, measuring the unexplained variance in the response that can be attributed to the nested structure of students, implies the existence of a substantial heterogeneity in the achievements among various classes. This highlight once again the importance of considering the hierarchical structure of these data.

Figure 6 reports the estimated random intercepts, that in this case are interpreted as school-effects, of OMERF and CLMM, that result to be coherent. Their distribution around 0, together with the high ICC, confirms that the likelihood of a student to be in different proficiency levels, net to the effect of his/her personal characteristics, is influenced by the attended school.



(a) ESCS (b) EMOCOAGR (c) CURIOAGR

(d) FAMSUP (e) ASSERAGR (f) gender

(g) STRESAGR (h) COOPAGR (i) HISCED

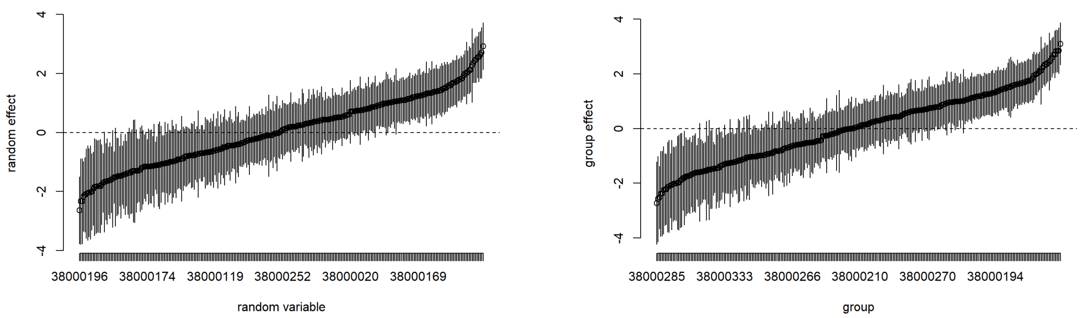
(j) PERSEVAGR (k) internet_quality (l) EMPATAGR

(m) BELONG (n) SCHRISK (o) BULLIED

(p) study_time (q) TEACHSUP (r) video_games

(s) grade (t) internet_availability (u) immig

Figure 5: Partial Plots for the fixed component of OMERF, in the real-world case study. The y-axis reports the increment/decrement of the target variable of the random forest in the iterative procedure, given the covariate on the x-axis.



(a) OMERF methos (b) CLMM method

Figure 6: Estimated Random intercepts with their 95% confidence intervals

6 Conclusions

In this study, we propose the Ordinal Mixed-Effects Random Forest (OMERF), a novel method that expands the utility of random forests to analyze hierarchical data with ordinal response variables. OMERF adopts the framework of cumulative linear mixed models but employs a random forest to estimate the fixed component. By doing so, it inherits the flexibility and predictive power of random forests while preserving the structure of mixed-effects models for ordinal response. This novel approach makes a valuable contribution to two branches of statistical literature: the one on tree-based mixed-effects models and the one on ordinal models.

A simulation study shows that OMERF outperforms the counterpart methods when the fixed component exhibits complex non linear structures, in presence of both light and strong heterogeneity at the group level, while linear models still perform better in case of linear predictors.

When applied to a real-world case study to predict mathematics proficiency levels of Italian students in the PISA 2022 test, OMERF performs similarly but slightly worse than CLMM, in terms of predictive power. Indeed, the strong linear relationship of the most important predictor (ESCS) with the response favors the linear model. Nonetheless, OMERF offers the advantage of investigating potential interactions and non linearities of other covariates, still disentangling the school effect on student outcomes, that, in the educational data mining context, as in many other contexts, is essential to properly understand students and the settings in which they learn.

In conclusion, OMERF proves to be a powerful and an easily interpretable method, that can deal with grouped data structures and can be applied to face complex real data challenges. Future developments might consider to enrich the modeling of the fixed-effect part by incorporating both a linear component and a tree-based component (see, for example, Gottard et al. (2019)). Moreover, OMERF should be employed to deal with more complex real-world data to test its efficacy in absence of clear linear predictors.

Acknowledgements: The present research is part of the activities of "Dipartimento di Eccellenza 2023-2027".

References

Agresti, A. (2010). *Analysis of ordinal categorical data*, volume 656. John Wiley & Sons.

Ananth, C. V. and Kleinbaum, D. G. (1997). Regression models for ordinal responses: a review of methods and applications. *International journal of epidemiology*, 26(6):1323–1333.

Ballante, E., Figini, S., and Uberti, P. (2022). A new approach in model selection for ordinal target variables. *Computational Statistics*, 37(1):43–56.

Breiman, L. (2001). Random forests. *Machine learning*, 45:5–32.

Breiman, L. (2017). *Classification and regression trees*. Routledge.

Cardoso, J. S. and Sousa, R. (2011). Measuring the performance of ordinal classification. *International Journal of Pattern Recognition and Artificial Intelligence*, 25(08):1173–1195.

Christensen, R. H. B. (2022). ordinal—regression models for ordinal data. R package version 2022.11-16. https://CRAN.R-project.org/package=ordinal.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1):37–46.

de Raadt, A., Warrens, M. J., Bosker, R. J., and Kiers, H. A. (2021). A comparison of reliability coefficients for ordinal rating scales. *Journal of Classification*, pages 1–25.

Fokkema, M., Smits, N., Zeileis, A., Hothorn, T., and Kelderman, H. (2018). Detecting treatment-subgroup interactions in clustered data with generalized linear mixed-effects model trees. *Behavior research methods*, 50:2016–2034.

Fontana, L., Masci, C., Ieva, F., and Paganoni, A. M. (2021). Performing learning analytics via generalised mixed-effects trees. *Data*, 6(7):74.

Gaudette, L. and Japkowicz, N. (2009). Evaluation methods for ordinal classification. In *Advances in Artificial Intelligence: 22nd Canadian Conference on Artificial Intelligence, Canadian AI 2009 Kelowna, Canada, May 25-27, 2009 Proceedings 22*, pages 207–210. Springer.

Goldfeld, K. and Wujciak-Jens, J. (2020). simstudy: Illuminating research methods through data generation. *Journal of Open Source Software*, 5(54):2763.

Gottard, A., Grilli, L., Rampichini, C., Vannucci, G., et al. (2019). Tree embedded linear mixed models. *Book of Short Papers*, page 239.

Grilli, L. and Rampichini, C. (2011). Multilevel models for ordinal data. *Modern analysis of customer surveys: With applications using R*, pages 391–411.

Hajjem, A., Bellavance, F., and Larocque, D. (2011). Mixed effects regression trees for clustered data. *Statistics & probability letters*, 81(4):451–459.

Hajjem, A., Bellavance, F., and Larocque, D. (2014). Mixed-effects random forest for clustered data. *Journal of Statistical Computation and Simulation*, 84(6):1313–1328.

Hajjem, A., Larocque, D., and Bellavance, F. (2017). Generalized mixed effects regression trees. *Statistics & Probability Letters*, 126:114–118.

Hornung, R. (2020). Ordinal forests. *Journal of Classification*, 37:4–17.

Hubert, L. and Arabie, P. (1985). Comparing partitions. *Journal of classification*, 2:193–218.

James, G., Witten, D., Hastie, T., Tibshirani, R., et al. (2013). *An introduction to statistical learning*, volume 112. Springer.

Liaw, A. and Wiener, M. (2002). Classification and regression by randomforest. *R News*, 2(3):18–22.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 42(2):109–127.

Pellagatti, M., Masci, C., Ieva, F., and Paganoni, A. M. (2021). Generalized mixed-effects random forest: A flexible approach to predict university student dropout. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 14(3):241–257.

Pinheiro, J. and Bates, D. (2006). *Mixed-effects models in S and S-PLUS*. Springer science & business media.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Raudenbush, S. W. and Willms, J. (1995). The estimation of school effects. *Journal of educational and behavioral statistics*, 20(4):307–335.

Sela, R. J. and Simonoff, J. S. (2012). Re-em trees: a data mining approach for longitudinal and clustered data. *Machine learning*, 86:169–207.

Shashua, A. and Levin, A. (2002). Ranking with large margin principle: Two approaches. *Advances in neural information processing systems*, 15.

Speiser, J. L., Wolf, B. J., Chung, D., Karvellas, C. J., Koch, D. G., and Durkalski, V. L. (2020). Bimm tree: a decision tree method for modeling clustered and longitudinal binary outcomes. *Communications in Statistics-Simulation and Computation*, 49(4):1004–1023.

Tutz, G. (2003). Generalized semiparametrically structured ordinal models. *Biometrics*, 59(2):263–273.

Tutz, G. (2022). Ordinal trees and random forests: Score-free recursive partitioning and improved ensembles. *Journal of Classification*, 39(2):241–263.

Tutz, G. and Hennevogl, W. (1996). Random effects in ordinal regression models. *Computational Statistics & Data Analysis*, 22(5):537–557.

Yang, Y., Chen, B., and Yang, Z. (2020). An algorithm for ordinal classification based on pairwise comparison. *Journal of Classification*, 37:158–179.

MOX Technical Reports, last issues

Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

40/2024 Carrara, D.; Regazzoni, F.; Pagani, S.
Implicit neural field reconstruction on complex shapes from scattered and noisy data

39/2024 Bartsch, J.; Buchwald, S.; Ciaramella, G.; Volkwein, S.
Reconstruction of unknown nonlienar operators in semilinear elliptic models using optimal inputs

38/2024 Tonini, A., Regazzoni, F., Salvador, M., Dede', L., Scrofani, R., Fusini, L., Cogliati, C., Pontone, G., Vergara, C., Quarteroni, A.
Two new calibration techniques of lumped-parameter mathematical models for the cardiovascular system

Fumagalli, A.; Patacchini, F.S.
Numerical validation of an adaptive model for the determination of nonlinear-flow regions in highly heterogeneous porous media

37/2024 Begu, B.; Panzeri, S.; Arnone, E.; Carey, M.; Sangalli, L.M.
A nonparametric penalized likelihood approach to density estimation of space-time point patterns

36/2024 Torri, V.; Ercolanoni, M.; Bortolan, F.; Leoni, O.; Ieva, F.
A NLP-based semi-automatic identification system for delays in follow-up examinations: an Italian case study on clinical referrals

34/2024 Corti, M.
Exploring tau protein and amyloid-beta propagation: a sensitivity analysis of mathematical models based on biological data

35/2024 Botti, L.; Botti, M.; Di Pietro, D.A.; Massa; F.C.
Stability, convergence, and pressure-robustness of numerical schemes for incompressible flows with hybrid velocity and pressure

Corti, M.
Exploring tau protein and amyloid-beta propagation: a sensitivity analysis of mathematical models based on biological data

32/2024 Ziarelli, G.; Parolini, N.; Verani, M.
Learning epidemic trajectories through Kernel Operator Learning: from modelling to optimal control