Modified Ordered Random Forest[*]

Riccardo Di Francesco[†]

April 26, 2023

Click here for the most recent version.

Abstract

Empirical studies in various social sciences often involve categorical outcomes with inherent ordering, such as self-evaluations of subjective well-being and self-assessments in health domains. While ordered choice models, such as the ordered logit and ordered probit, are popular tools for analyzing such outcomes, they may impose restrictive parametric and distributional assumptions. This paper provides a novel estimator, the *modified ordered random forest,* which overcomes these limitations. The proposed estimator modifies a standard random forest splitting criterion to build a collection of forests, each estimating the conditional probability of a single class. Under an "honesty" condition, predictions are consistent and asymptotically normal. The weights induced by each forest are used to estimate the variance of the predictions and obtain estimation and inference about the covariates' marginal effects. Evidence from synthetic and real data sets shows that the proposed estimator features a superior prediction performance than alternative estimators.

Keywords: Ordered non-numeric outcomes, choice probabilities, machine learning.
JEL Codes: C14, C25, C55

[*] I especially would like to thank Franco Peracchi for feedback and suggestions. I am also grateful to Michael Lechner, Jana Mareckova, and seminar participants at SEW-HSG research seminars for comments and discussions. Gabriel Okasa generously shared the code for implementing the DGPs in the simulation. The R package for implementing the methodology developed in this paper is available at https://github.com/riccardo-df/morf. The associated vignette is at https://riccardo-df.github.io/morf/.

[†] Department of Economics and Finance, University of Rome Tor Vergata, Rome. Electronic correspondence: riccardo.di.francesco@uniroma2.it.

1 Introduction

Categorical outcomes with a natural order are commonly observed in empirical studies across social sciences. For example, happiness research typically employs large surveys to collect self-evaluations of subjective well-being (Frey & Stutzer, 2002), and health economics is heavily based on self-assessments in several health domains (see, e.g., Peracchi & Rossetti, 2012, 2013). These outcomes are usually measured on a discrete scale with five or ten classes, where the classes can be arranged in a natural order without any knowledge about their relative magnitude.

Ordered choice models are a popular class of statistical models used to analyze the relationship between this kind of outcome and a set of covariates (see e.g., Greene & Hensher, 2010). These models target the estimation of the conditional choice probabilities, which represent the probability that the outcome belongs to a certain class given the values of the covariates. Common examples of ordered choice models include ordered probit and ordered logit models. However, these models are limited by their dependence on parametric and distributional assumptions that are often based on analytical convenience rather than knowledge about the underlying data generating process. As a result, econometricians may need to consider alternative techniques to produce more accurate and reliable predictions.

Recent developments in statistical learning offer new ways to relax these assumptions. Nonparametric estimators, such as decision trees (Breiman et al., 1984), random forests (Breiman, 2001), and boosting (Friedman, 2001), have been developed to accommodate continuous or discrete outcomes. For ordered non-numeric (i.e., categorical) outcomes, one possible adaptation involves expressing conditional probabilities as the difference between the cumulative probabilities of two adjacent classes, which transforms the problem into estimating the difference between two conditional expectations. Then, any machine learning algorithm can be used to estimate each expectation individually, and the difference between the estimated surfaces can be used to recover the conditional probabilities. For example, Lechner and Okasa (2019) combine this strategy with random forests.

However, this estimation strategy suffers from two main limitations. First, it fails to account for the potential correlation between the estimation errors of the cumulative probabilities. Correlating the errors could improve estimation performance since errors that move in the same direction cancel out when taking the difference between the estimated surfaces. Second, it may produce negative predictions, which contradicts the definition of probabilities.

This paper provides a novel estimator that addresses these limitations, the *modified ordered random forest*. The proposed estimator modifies a standard random forest splitting criterion to build a collection of forests, each estimating the conditional probability of a single class. Intuitively, each forest ties the estimation of the cumulative probabilities of two adjacent classes to correlate the estimation errors. This is achieved by using a splitting rule that penalizes splits that would induce a low or negative correlation. Once the forests are built, an unbiased estimator of conditional probabilities is used in each leaf. Model consistency is ensured, as the predictions always lie in the unit interval by construction. Evidence from synthetic and real data sets shows that the modified ordered random forest features a superior prediction performance than alternative estimators.

The proposed estimator inherits the asymptotic properties of random forests proven by Wager and Athey (2018), namely the consistency and asymptotic normality of its predictions. This allows valid inference about conditional probabilities to be made using conventional methods, although requires the individual trees to satisfy a fairly strong condition called honesty (Athey & Imbens, 2016). Honesty is a subsample-splitting technique that ensures that different observations are used to place the splits and compute leaf predictions and is crucial to achieving consistency of the predictions. The particular honesty implementation used by the modified ordered random forest estimator allows for a weight-based estimation of the variance of the predicted probabilities. This is achieved by rewriting the random forest predictions as a weighted average of the outcomes (Athey et al., 2019). The weights, which are obtained for the predicted probabilities, can be properly transformed to obtain estimation and inference about the covariates' marginal effects (for a similar approach, see

Lechner & Okasa, 2019; Lechner & Mareckova, 2022).

The rest of the paper unfolds as follows. Section 2 provides a brief overview of the ordered choice models and discusses some alternatives proposed in the literature. Section 3 presents the modified ordered random forest estimator, explaining estimation and inference about the statistical targets of interest. Section 4 uses synthetic and real data sets to compare the modified ordered random forest with alternative estimators. Section 5 concludes.

2 Ordered Choice Models

Ordered choice models are a class of statistical models used to analyze the relationship between an ordered non-numeric outcome Y_i and a set of covariates X_i (McCullagh, 1980). These models are typically motivated by postulating the existence of a latent and continuous outcome variable of interest Y_i^*, assumed to be linearly related to the covariates through unknown coefficients β and subject to random error U_i:

$$Y_i^* = X_i^T \beta + U_i, \quad U_i | X_i \sim f(0, \sigma^2) \tag{2.1}$$

However, we observe only the discretized version Y_i of Y_i^*, which takes on integer values $m = 1, \dots, M$ corresponding to different categories or classes. Unknown threshold parameters $-\infty = \zeta_0 < \zeta_1 < \cdots < \zeta_{M-1} < \zeta_M = \infty$ define intervals on the support of Y_i^*, each corresponding to one of the M categories of the observed variable Y_i:

$$\zeta_{m-1} < Y_i^* \leq \zeta_m \implies Y_i = m, \quad m = 1, \dots, M \tag{2.2}$$

Although the M classes have a natural ordering, their relative magnitude is unknown, thus limiting our ability to make precise quantitative comparisons.

Researchers are typically interested in the estimation of the conditional choice probabilities, defined as:

$$p_m(X_i) := \mathbb{P}(Y_i = m | X_i) \tag{2.3}$$

However, the marginal effect of the j-th covariate on $p_m(\cdot)$ is a more interpretable measure for ordered choice models. The marginal effect is defined based on the continuous or discrete nature of the covariate:

$$p'_{m,j}(x) := \frac{\partial p_m(x)}{\partial x_j} \tag{2.4}$$

$$p'_{m,j}(x) := p_m(\lceil x_j \rceil) - p_m(\lfloor x_j \rfloor) \tag{2.5}$$

where x_j is the j-th element of the vector x and $\lceil x_j \rceil$ and $\lfloor x_j \rfloor$ correspond to x with its j-th element rounded up and down to the closest integer. We can summarize the marginal effects in various ways, such as computing the marginal effect at the mean $p'_{m,j}(\bar{x})$, with $\bar{x}$ denoting a vector of means. Alternatively, we can compute the marginal effect at the median, the mean marginal effect, and the median marginal effect.

Assumptions on the error term distribution f are generally imposed to derive a closed-form expression of the conditional probabilities. Popular choices of f are the standard normal distribution function and the standard logistic distribution function, which produce the ordered probit and ordered logit models. Estimation is generally performed using standard maximum likelihood methods.

Although easy to interpret and computationally efficient, these models feature several limitations. First, they impose strong distributional and functional form assumptions generally derived from analytical convenience rather than knowledge about the underlying data generating process. Second, the definition and estimation of the marginal effects have the restrictive property of single-crossing, meaning that these effects can change sign only once when moving from the smallest class to the largest. Third, if the number of covariates is larger than the number of observations, estimation breaks down.

Several alternatives have been proposed in the literature to overcome these limitations. Boes and Winkelmann (2006) discuss generalizations of the standard ordered choice models, but they still rely on parametric and distributional assumptions, limiting the increase in flexibility they provide.

Machine learning techniques, such as decision trees (Breiman et al., 1984), allow for a more flexible analysis of ordered categorical outcomes. Piccarreta (2008) discusses several criteria for constructing classification trees that account for the ordered structure of the outcome. However, trees exhibit a large sampling variance and aim to estimate only outcome categories, rather than conditional probabilities.

Other estimation strategies are based on the random forest algorithm introduced by Breiman (2001). Standard classification and regression forests are often used to analyze ordered categorical outcomes. However, classification forests do not leverage the ordering information, and regression forests treat the outcome as if it is measured on a metric scale.

Hothorn et al. (2006) propose to transform ordered non-numeric outcomes to a metric scale using scores $s(1), \ldots, s(M)$ constructed based on the classes $1, \ldots, M$ of the observed outcome Y_i. These scores can then be used as an outcome in any machine learning algorithm, such as regression forest. Hothorn et al. (2006) suggest using the midpoints of the intervals defined on the support of the latent outcome Y_i^* as score values. Because Y_i^* is generally not observed, this translates into setting the scores equal to the class labels, that is, $s(m) = m$.

Hornung (2020) notes that setting $s(m) = m$ implicitly assumes that the intervals defined on the support of Y_i^* are of equal length, which may not be true. To address this limitation, he proposes the ordinal forest estimator, which optimizes the class intervals and uses score values that correspond to these optimized intervals in a standard regression forest. The optimization involves growing multiple forests using randomly generated candidate score sets and constructing the final score values as a summary of the score sets that featured the smallest out-of-bag error. Hornung (2020) uses both real and synthetic data sets to show that the ordinal forest estimator performs particularly well compared to a standard regression forest that uses the class labels as score values. However, the optimization process greatly increases computational time, which may make the estimator impractical for large data sets or real-time applications.

Finally, Lechner and Okasa (2019) introduce the ordered random forest estimator, which, in an extensive simulation study, outperforms conditional forests (Hothorn et al., 2006) and ordinal forests (Hornung, 2020) in the most complex designs. This estimator constructs separate regression forests to estimate the cumulative probability of each class and takes the difference between the cumulative probabilities of two adjacent classes m and $m-1$ to estimate $p_m(\cdot)$. However, this estimation strategy suffers from two main limitations. First, it fails to account for the potential correlation between the estimation errors of the cumulative probabilities. Correlating the errors could improve estimation performance since errors that move in the same direction cancel out when taking the difference between the estimated surfaces. Second, it may produce negative predictions, which contradicts the definition of probabilities. The methodology proposed in this paper, the modified ordered random forest, modifies the standard random forest algorithm to address these limitations.

Lechner and Okasa (2019) also propose a nonparametric estimator of the covariates' marginal effects that approximates the infinitesimal change in x_j via its discrete counterpart and leverage the weights induced by the forests (Athey et al., 2019) to estimate standard errors (see Lechner & Mareckova, 2022, for a similar approach). The modified ordered random forest uses these ideas to obtain estimation and inference about the marginal effects.

3 Estimation and Inference

This section introduces the modified ordered random forest ($MORF$) estimator, which aims to overcome the limitations of the ordered random forest (ORF) estimator proposed by Lechner and Okasa (2019). Both estimators construct a collection of random forests to estimate conditional probabilities. However, while ORF relies on the standard implementation (Breiman, 2001), $MORF$ modifies the algorithm by changing how individual trees are constructed and how predictions are computed.

In the following subsection, I provide an overview of the ORF estimator and highlight its main limitations. Next, I introduce the $MORF$ estimator and explain how it addresses these limitations. I then discuss the conditions necessary for the asymptotic normality and consistency of $MORF$ predictions. Finally, I demonstrate how to conduct approximate inference about the statistical targets of interest.

3.1 Ordered Random Forest

Lechner and Okasa (2019) notice that conditional choice probabilities can be expressed as the difference between the cumulative probabilities of two adjacent classes:

$$\begin{aligned} p_m(X_i) &= \mathbb{P}(Y_i \leq m | X_i) - \mathbb{P}(Y_i \leq m-1 | X_i) \\ &= \mathbb{E}[\mathbb{1}(Y_i \leq m) | X_i] - \mathbb{E}[\mathbb{1}(Y_i \leq m-1) | X_i] \\ &= \mu_m(X_i) - \mu_{m-1}(X_i) \end{aligned} \tag{3.1}$$

Thus they propose the ORF estimator, which separately estimates $\mu_m(\cdot)$ for all $m = 1, \ldots, M-1$ using standard regression forests and picks the difference between the cumulative probabilities of two adjacent classes to estimate $p_m(\cdot)$:[1]

$$\hat{p}_m^{ORF}(X_i) = \hat{\mu}_m(X_i) - \hat{\mu}_{m-1}(X_i) \tag{3.2}$$

However, ORF suffers from two main limitations. First, it fails to account for the potential correlation between the errors made in estimating $\mu_m(\cdot)$ and $\mu_{m-1}(\cdot)$. This can be shown by decomposing the mean squared error of a prediction $\hat{p}_m^{ORF}(\cdot)$ at x as follows:[2]

$$\begin{aligned} MSE\left(\hat{p}_m^{ORF}(x)\right) &= \mathbb{E}\left[\left\{\hat{p}_m^{ORF}(x) - p_m(x)\right\}^2\right] \\ &= \mathbb{E}\left[\{\hat{\mu}_m(x) - \hat{\mu}_{m-1}(x) - \mu_m(x) + \mu_{m-1}(x)\}^2\right] \\ &= MSE(\hat{\mu}_m(x)) + MSE(\hat{\mu}_{m-1}(x)) - 2EC(\hat{\mu}_m(x), \hat{\mu}_{m-1}(x)) \end{aligned} \tag{3.3}$$

[1] Estimation of the last cumulative probability is not needed as $\mu_M(x) = 1$ for all x by construction.

[2] This decomposition can be applied to any estimation strategy that involves calculating the difference between two surfaces. For example, Lechner and Mareckova (2022) leverage this decomposition to estimate heterogeneous causal effects.

where the last term is the error correlation and captures the degree to which the errors made in estimating $\mu_m(\cdot)$ and $\mu_{m-1}(\cdot)$ are correlated:

$$EC\left(\hat{\mu}_m(x),\, \hat{\mu}_{m-1}(x)\right) = \mathbb{E}\left[\left\{\hat{\mu}_m(x) - \mu_m(x)\right\}\left\{\hat{\mu}_{m-1}(x) - \mu_{m-1}(x)\right\}\right] \tag{3.4}$$

Equation (3.3) shows that ORF is a suboptimal estimator. Estimating $\mu_m(\cdot)$ and $\mu_{m-1}(\cdot)$ separately minimizes only the mean squared error terms and ignores the error correlation. Correlating the errors could improve estimation performance since errors that move in the same direction cancel out when taking the difference $\hat{\mu}_m(\cdot) - \hat{\mu}_{m-1}(\cdot)$.

Second, ORF may produce negative predictions, which contradicts the definition of probabilities. Although Lechner and Okasa (2019) resolve this issue by setting negative predictions to zero, an alternative estimator that does not require truncation may perform better.

3.2 Modified Ordered Random Forest

The estimator proposed in this paper, $MORF$, modifies the standard random forest algorithm to address the limitations of ORF. In particular, $MORF$ constructs individual trees using a splitting rule that ties the estimation of $\mu_m(\cdot)$ and $\mu_{m-1}(\cdot)$, and computes predictions that always lie in the unit interval.

Similar to ORF, $MORF$ constructs a collection of forests, one for each of the M classes of Y_i. As in the standard algorithm, individual trees are constructed by recursively partitioning the covariate space using axis-aligned splits. Each split is chosen to partition a *"parent node"* $\mathcal{P} \subseteq \mathcal{X}$ into two *"child nodes"* $C_1, C_2 \subset \mathcal{P}$ such as to minimize the assumed loss function as much as possible. The process is then repeated in the resulting nodes until some stopping criterion is met, e.g., a minimum number of observations in the *"leaves"* (i.e., terminal nodes) of the tree.

However, rather than the standard criterion, $MORF$ uses equation (3.3) as the splitting rule to build the individual trees in the m-th forest. This allows us to tie the estimation of $\mu_m(\cdot)$ and $\mu_{m-1}(\cdot)$, thus accounting for the error correlation that ORF ignores. Intuitively,

$MORF$ penalizes splits that would induce a low or negative correlation between the errors made in estimating $\mu_m(\cdot)$ and $\mu_{m-1}(\cdot)$.

To use (3.3) as the splitting rule, we need to estimate its components. This, in turn, requires an estimator of $\mu_m(\cdot)$ in each node. An unbiased estimator of $\mu_m(\cdot)$ in a child node $C_j \subset \mathcal{P}$ consists of the proportion of observations in C_j whose outcome is lower than or equal to m:

$$\check{\mu}_m(X_i) = \frac{1}{|C_j|} \sum_{i:X_i \in C_j} \mathbb{1}(Y_i \leq m) \tag{3.5}$$

This leads us to estimating $MSE(\check{\mu}_m(\cdot))$ and $EC(\check{\mu}_m(\cdot), \check{\mu}_{m-1}(\cdot))$ in each node by their sample analogs:

$$\widehat{MSE}_{C_j}(\check{\mu}_m(X_i)) = \frac{1}{|C_j|} \sum_{i:X_i \in C_j} \left[\mathbb{1}(Y_i \leq m) - \check{\mu}_m(X_i)\right]^2 \tag{3.6}$$

$$\widehat{EC}_{C_j}(\check{\mu}_m(X_i), \check{\mu}_{m-1}(X_i)) = \frac{1}{|C_j|} \sum_{i:X_i \in C_j} \mathbb{1}(Y_i \leq m)\, \mathbb{1}(Y_i \leq m-1) - \check{\mu}_m(X_i)\, \check{\mu}_{m-1}(X_i) \tag{3.7}$$

Then, in the m-th forest, $MORF$ constructs individual trees by recursively partitioning each parent node into two child nodes C_1 and C_2 such that the following minimization problem is solved:

$$\min_{C_1, C_2} \sum_{j=1}^{2} \widehat{MSE}_{C_j}(\check{\mu}_m(X_i)) + \widehat{MSE}_{C_j}(\check{\mu}_{m-1}(X_i)) - 2\widehat{EC}_{C_j}(\check{\mu}_m(X_i), \check{\mu}_{m-1}(X_i)) \tag{3.8}$$

Once the recursive partitioning stops, each tree in the m-th forest unbiasedly estimates $p_m(\cdot)$ at x by computing the proportion of observations in the same leaf as x whose outcome equals m:

$$\begin{aligned} \hat{p}_{m,b}^{MORF}(x) &= \check{\mu}_m(x) - \check{\mu}_{m-1}(x) \\ &= \frac{1}{|L_{m,b}(x)|} \sum_{i \in L_{m,b}(x)} \mathbb{1}(Y_i = m) \end{aligned} \tag{3.9}$$

where $L_{m,b}(x)$ is the set of observations falling in the same leaf of the b-th tree as the prediction point x. The predictions from each tree are then averaged to obtain the forest predictions:[3]

$$\hat{p}_m^{MORF}(x) = \frac{1}{B_m} \sum_{b=1}^{B_m} \hat{p}_{m,b}^{MORF}(x) \tag{3.10}$$

where $b = 1, \ldots, B_m$ indexes the trees in the m-th forest. In contrast to ORF, $MORF$ ensures model consistency, as the predictions $\hat{p}_m^{MORF}(\cdot)$ always lie in the unit interval by construction.

Estimation of marginal effects proceeds as proposed by Lechner and Okasa (2019). For discrete covariates, we can plug an estimate $\hat{p}_m^{MORF}(\cdot)$ of $p_m(\cdot)$ into equation (2.5) to have a straightforward estimator of $p'_{m,j}(\cdot)$. For continuous covariates, we use a nonparametric approximation of the infinitesimal change in x_j:

$$\hat{p}'^{MORF}_{m,b}(x) = \frac{\hat{p}_m^{MORF}(\widehat{\lceil x_j \rceil}) - \hat{p}_m^{MORF}(\widehat{\lfloor x_j \rfloor})}{\bar{x}_j - \underline{x}_j} \tag{3.11}$$

where $\widehat{\lceil x_j \rceil}$ and $\widehat{\lfloor x_j \rfloor}$ correspond to x with its j-th element set to $\bar{x}_j = x_j + \omega \sigma_j$ and $\underline{x}_j = x_j - \omega \sigma_j$, with σ_j the standard deviation of x_j and $\omega > 0$ a tuning parameter.

3.3 Asymptotic Properties

Wager and Athey (2018) establish the consistency and asymptotic normality of random forest predictions. However, besides some regularity and technical assumptions, there are certain conditions regarding the construction of individual trees that must be satisfied. In the following, I define these conditions.

The first condition requires that the trees use different observations to place the splits and compute the leaf predictions. This condition is called *honesty* and is crucial to bounding the bias of forest predictions.

Definition 1 (*Honesty*)**.** *A tree is honest if it uses the outcome Y_i to either place the splits or compute the leaf predictions, but not both.*

[3] To ensure that $\sum_{m=1}^{M} \hat{p}_m^{MORF}(x) = 1$, a normalization step may be necessary.

Wager and Athey (2018) implement honesty by first drawing a subsample from the original sample $\mathcal{S}$ for each tree and then splitting the subsample into two halves, using one half to grow the tree and the other half to compute leaf predictions (see also Athey et al., 2019). Alternatively, Lechner and Mareckova (2022) suggest a different implementation (also used by Lechner & Okasa, 2019). First, they split the original sample $\mathcal{S}$ into a training sample $\mathcal{S}^{tr}$ and an honest sample $\mathcal{S}^{hon}$. Then, they use random subsamples from $\mathcal{S}^{tr}$ to construct trees and compute leaf predictions using only $\mathcal{S}^{hon}$. This strategy enables weight-based inference about leaf predictions and their transformations, such as marginal effects. $\boldsymbol{MORF}$ adopts this strategy as well (see Section 3.4).

The second condition is that the leaves of the trees must become small in all dimensions of the covariate space as the sample size increases. This is necessary for achieving consistency of the predictions and is accomplished by introducing randomness in the tree-growing process and enforcing a regularity condition on how quickly the leaves get small.

Definition 2 (*Random-split*)**.** *A tree is random-split if, at every step of the tree-growing procedure, the probability that the next split occurs along the* $j-th$ *covariate is bounded below by* π/k*, for some* $0 < \pi \leq 1$*, for all* $j = 1, \dots, k$*.*

Definition 3 (*α-regularity*)**.** *A tree is* α*-regular if each split leaves at least a fraction* α *of the observations in the parent node on each side of the split and the trees are fully grown to depth* d *for some* $d \in N$*, that is, there are between* d *and* $2d-1$ *observations in each terminal node of the tree.*

To achieve α-regularity, $\boldsymbol{MORF}$ ignores splits that do not satisfy this condition. The algorithm always selects the best split from among the candidate splits that would maintain at least a fraction α of the parent node's observations on both sides of the split. This way, we can rule out any influence of the splitting rule on the shape of the final leaves.

Third, trees must be constructed using subsamples drawn without replacement, rather than bootstrap samples, as originally proposed by Breiman (2001).

Lastly, to derive the asymptotic normality, trees must be symmetric.

Definition 4 (*Symmetry*)**.** *A predictor is symmetric if the (possibly randomized) output of the predictor does not depend on the order in which observations are indexed in the training and honest samples.*

Under these conditions, we can establish consistency and asymptotic normality of the predictions. For completeness, I here report the main theorem by Wager and Athey (2018).

Theorem 3.1. *Suppose that we have n independent and identically distributed training examples $(X_i, Y_i) \in [0,1]^k \times \mathcal{R}$. Suppose moreover that the covariates are independently and uniformly distributed $X_i \sim U([0,1]^k)$, that $\theta(x) = \mathbb{E}[Y|X=x]$ and $\theta_2(x) = \mathbb{E}\left[Y^2|X=x\right]$ are Lipschitz-continuous, and finally that $Var[Y|X=x] > 0$ and $\mathbb{E}\left[|Y - \mathbb{E}[Y|X=x]|^{2+\delta}|X=x\right] \leq M$ for some constants $\delta, M > 0$, uniformly over all $x \in [0,1]^k$. Given this data-generating process, let $\mathcal{T}$ be an honest, α-regular with $\alpha \leq 0.2$, and symmetric random-split tree in the sense of Definitions 1–4, and let $\hat{\theta}_n(x)$ be the estimate for $\theta(x)$ given by a random forest with base learner $\mathcal{T}$ and a subsample size s_n. Finally, suppose that the subsample size s_n scales as:*

$$s_n \asymp n^{\beta} \text{ for some } \beta_{min} := 1 - \left(1 + \frac{k}{\pi}\frac{log\left(\alpha^{-1}\right)}{log\left(\left(1-\alpha\right)^{-1}\right)}\right) < \beta < 1$$

Then, random forest predictions are asymptotically Gaussian:

$$\frac{\hat{\theta}_n(x) - \theta(x)}{\sigma_n(x)} \Longrightarrow \mathcal{N}(0,1) \text{ for a sequence } \sigma_n(x) \to 0$$

3.4 Inference

In addition to Theorem 3.1, Wager and Athey (2018) show that the asymptotic variance of random forest predictions $\sigma_n(\cdot)$ can be consistently estimated by adapting the infinitesimal jackknife estimator proposed by Wager et al. (2014) to the case of subsampling without replacement. This approach can be used to estimate the variance of a prediction $\hat{p}_m^{MORF}(\cdot)$ at

x. However, generalizing this method to estimate the variance of marginal effects $\hat{p}'^{MORF}_{m,j}(\cdot)$ is not straightforward.

To overcome this limitation, $MORF$ employs an alternative approach that leverages the weight-based representation of random forest predictions (Athey et al., 2019) and adapts the weight-based inference proposed by Lechner and Mareckova (2022) (see also Lechner & Okasa, 2019). In particular, $MORF$ implements honesty in a way that guarantees that the weight assigned to the i-th unit is independent of the outcomes of other units. This allows for the derivation of a straightforward formula for the variance of honest predicted probabilities and marginal effects.

First, we express $MORF$ predictions as weighted averages of the outcomes. Let $\mathcal{S}$ denote the observed sample. The following provides an expression for a prediction $\hat{p}^{MORF}_{m}(\cdot)$ at x numerically equivalent to that in (3.10):

$$\begin{gathered}\hat{p}^{MORF}_{m}(x) = \sum_{i \in \mathcal{S}} \hat{\alpha}_{m,i}(x) \, \mathbb{1}(Y_i = m) \\ \hat{\alpha}_{m,b,i}(x) = \frac{\mathbb{1}\left(X_i \in L_{m,b}(x)\right)}{\left|L_{m,b}(x)\right|}, \quad \hat{\alpha}_{m,i}(x) = \frac{1}{B_m} \sum_{b=1}^{B_m} \hat{\alpha}_{m,b,i}(x)\end{gathered} \tag{3.12}$$

where the weights $\hat{\alpha}_{m,1}(x), \ldots, \hat{\alpha}_{m,|\mathcal{S}|}(x)$ determine the forest-based adaptive neighborhood of x. They represent how often the i-th observation in $\mathcal{S}$ shares a leaf with x in the m-th forest. This measures how important the i-th observation is for fitting $p_m(\cdot)$ at x. Notice that $\sum_{i \in \mathcal{S}} \hat{\alpha}_{m,i}(x) = 1$ for all x.

Calculating the variance of a prediction $\hat{p}^{MORF}_{m}(x)$ in (3.12) is challenging because the weight assigned to the i-th unit $\hat{\alpha}_{m,i}(x)$ is a function of both $\mathcal{S}$ and X_i. Thus, this weight depends on the outcomes of all other units in $\mathcal{S}$, which complicates the formula for the variance.

However, the formula for the variance simplifies under the particular honesty implementation of $MORF$. Let $\mathcal{S}^{tr}$ and $\mathcal{S}^{hon}$ be a training sample and an honest sample obtained by randomly splitting the observed sample $\mathcal{S}$. Also, let $\hat{\alpha}^{tr}_{m,i}(\cdot)$ be the weights induced by a forest constructed using only $\mathcal{S}^{tr}$. Then, an honest prediction $\hat{p}_m^{MORF_H}(\cdot)$ at x is obtained by the following weighted average of observations in $\mathcal{S}^{hon}$:

$$\hat{p}_m^{MORF_H}(x) = \sum_{i \in \mathcal{S}^{hon}} \hat{\alpha}_{m,i}^{tr}(x) \mathbb{1}(Y_i = m) \tag{3.13}$$

The new weight assigned to the i-th unit $\hat{\alpha}_{m,i}^{tr}(x)$ is a function of $\mathcal{S}^{tr}$ and of X_i. Thus, under i.i.d. sampling this weight is independent of the outcomes of other units in $\mathcal{S}^{hon}$. This allows us to derive a simple formula for the variance of an honest prediction $\hat{p}_m^{MORF_H}(x)$:

$$\mathbb{V}\left(\hat{p}_m^{MORF_H}(x)\right) = |\mathcal{S}^{hon}| \, \mathbb{V}\left(\hat{\alpha}_{m,i}^{tr}(x) \mathbb{1}(Y_i = m)\right) \tag{3.14}$$

We can estimate this variance by its sample analog.

By plugging (3.13) into (3.11), we obtain the following estimator of honest marginal effects:

$$\begin{aligned} \hat{p}'^{MORF_H}_{m,j}(x) &= \frac{1}{\bar{x}_j - \underline{x}_j} \left\{ \sum_{i \in \mathcal{S}^{hon}} \hat{\alpha}_{m,i}^{tr}(\widehat{\lceil x_j \rceil}) \mathbb{1}(Y_i = m) - \sum_{i \in \mathcal{S}^{hon}} \hat{\alpha}_{m,i}^{tr}(\widehat{\lfloor x_j \rfloor}) \mathbb{1}(Y_i = m) \right\} \\ &= \frac{1}{\bar{x}_j - \underline{x}_j} \sum_{i \in \mathcal{S}^{hon}} \tilde{\alpha}_{m,i}^{tr}(\widehat{\lceil x_j \rceil}, \widehat{\lfloor x_j \rfloor}) \mathbb{1}(Y_i = m) \end{aligned} \tag{3.15}$$

with $\tilde{\alpha}_{m,i}^{tr}(\widehat{\lceil x_j \rceil}, \widehat{\lfloor x_j \rfloor}) = \hat{\alpha}_{m,i}^{tr}(\widehat{\lceil x_j \rceil}) - \hat{\alpha}_{m,i}^{tr}(\widehat{\lfloor x_j \rfloor})$ a transformation of the original weights. Using the same argument as before, under i.i.d. sampling the weight assigned to the i-th unit $\tilde{\alpha}_{m,i}^{tr}(\widehat{\lceil x_j \rceil}, \widehat{\lfloor x_j \rfloor})$ is independent of the outcomes of other units in $\mathcal{S}^{hon}$. Thus the variance of an honest marginal effect $\hat{p}'^{MORF_H}_{m,j}(x)$ can be expressed as follows:

$$\mathbb{V}\left(\hat{p}'^{MORF_H}_{m,j}(x)\right) = \frac{|\mathcal{S}^{hon}|}{\left(\bar{x}_j - \underline{x}_j\right)^2} \mathbb{V}\left(\tilde{\alpha}_{m,i}^{H}(\widehat{\lceil x_j \rceil}, \widehat{\lfloor x_j \rfloor}) \mathbb{1}(Y_i = m)\right) \tag{3.16}$$

As before, we can estimate this variance by its sample analog.

According to Theorem 3.1, the honest predicted probabilities in (3.13) are consistent and asymptotically normal, provided that the weights $\hat{\alpha}_{m,i}^{tr}(\cdot)$ are induced by a forest composed of α-regular with $\alpha \leq 0.2$ and symmetric random-split trees grown using subsampling without replacement. With these conditions met, we can use the estimated standard errors of honest predicted probabilities $\hat{p}_m^{MORF_H}(\cdot)$ to conduct valid inference as usual, e.g., by constructing conventional confidence intervals.

Furthermore, under the same conditions the honest marginal effects in (3.15) are a linear combination of normally distributed predictions, and thus have a normal distribution as well. Therefore, we can also construct conventional confidence intervals for honest marginal effects $\hat{p}_{m,j}^{\prime,MORF_H}(\cdot)$ using their estimated standard errors.

4 Comparison with Alternative Estimators

In this section, I compare the modified ordered random forest ($MORF$) with the ordered random forest (ORF) estimator of Lechner and Okasa (2019) using synthetic and real data sets. I also consider the ordered logit (OL) model as a parametric benchmark for the comparison.

4.1 Simulation Results

I choose the DGPs to replicate part of the simulation study in Lechner and Okasa (2019). I consider three different designs that differ in the model for the latent outcome variable:[4]

$$\begin{aligned} &\textit{Design 1.} && Y_i^* = X_i^T\beta + U_i \\ &\textit{Design 2.} && Y_i^* = \sum_{j=1}^{k} X_{ij}\mathbb{1}\left(X_{ij} > 0\right)\beta_j + U_i \\ &\textit{Design 3.} && Y_i^* = sin\left(2X_i\right)^T\beta + U_i \end{aligned}$$

with $U_i \sim logistic\,(0,1)$ in all designs. The three designs share all the other settings described below. For each design, I consider four sample sizes, $|\mathcal{S}| \in \{500,\ 1000, 2000, 4000\}$. Thus, I consider overall twelve different scenarios.

[4] *Design 2* is not included in the study of Lechner and Okasa (2019).

I obtain the observed outcomes Y_i by discretizing Y_i^* into nine classes:

$$\zeta_{m-1} < Y_i^* \leq \zeta_m \implies Y_i = m, \;\; m = 1, \ldots 9$$

I construct the threshold parameters $\zeta_1, \ldots, \zeta_8$ as follows. First, I draw eight values $\zeta_m^q \sim U\left(0.09, 0.91\right)$ and sort them in ascending order, so that $\zeta_m^q \leq \zeta_{m+1}^q$.[5] Then, I generate a sample of 1,000,000 Y_i^* and set $\zeta_m = Q\left(\zeta_m^q\right)$, with $Q\left(\cdot\right)$ the empirical quantile function of Y_i^*. This way, the threshold parameters are unevenly spaced, and the class widths are randomized and unequal.

I generate $k = 30$ covariates $X_i \sim \mathcal{N}\left(0, \Sigma\right)$. The components of the coefficient vector β are $\beta_1, \ldots, \beta_5 = 1$, $\beta_6, \ldots, \beta_{10} = 0.75$, and $\beta_{11}, \ldots, \beta_{15} = 0.5$. The remaining covariates have null coefficients, that is, they are "noise" covariates. The variance-covariance matrix Σ is block diagonal and induces correlation among signal covariates as well as among noise covariates, but there is zero correlation between signal and noise covariates:

$$\Sigma = \begin{pmatrix} \mathbf{A}_{signal} & 0 \\ 0 & \mathbf{A}_{noise} \end{pmatrix} \qquad a_{i,j}^{signal} = a_{i,j}^{noise} = \begin{cases} 1, & i = j \\ 0.8, & i \neq j \cap \{i, j \; are \; odd\} \\ 0, & otherwise \end{cases}$$

After drawing a sample $\mathcal{S}$, I estimate the conditional choice probabilities using OL, ORF, and two versions of $MORF$, the "adaptive" version $MORF_A$ and the "honest" version $MORF_H$. This way, we can quantify the loss in the precision derived from using fewer observations to build the forests, representing the price to pay for valid inference. I feed OL with all covariates without adding any polynomials, interaction terms, or other transformations of the covariates. Thus, OL is correctly specified in *Design 1* and misspecified in the other designs. To implement $MORF_H$, I randomly split $\mathcal{S}$ into a training sample $\mathcal{S}^{tr}$ used to construct the trees and an honest sample $\mathcal{S}^{hon}$ used to compute the leaf predictions. I choose $|\mathcal{S}^{tr}| = |\mathcal{S}^{hon}| = |\mathcal{S}|/2$.

I rely on an external validation sample $\mathcal{S}^{val}$ of size $|\mathcal{S}^{val}| = 10,000$ to assess the predic-

[5] If the distance between two adjacent values ζ_m^q and ζ_{m+1}^q is not sufficient, I redraw another set of values.

tion performance of the estimators. This large number of observations helps minimize the sampling variance. For each replication $r = 1, \dots, R$, I calculate the mean squared error and ranked probability score for each estimator:

$$MSE_r = \frac{1}{|\mathcal{S}^{val}|} \sum_{i \in \mathcal{S}^{val}} \sum_{m=1}^{M} \left[p_m\left(X_i\right) - \hat{p}_{m,r}\left(X_i\right) \right]^2 \tag{4.1}$$

$$RPS_r = \frac{1}{|\mathcal{S}^{val}|} \sum_{i \in \mathcal{S}^{val}} \frac{1}{M-1} \sum_{m=1}^{M} \left[\mu_m\left(X_i\right) - \hat{\mu}_{m,r}\left(X_i\right) \right]^2 \tag{4.2}$$

with $\hat{p}_{m,r}(\cdot)$ the estimated conditional probabilities in the r-th replication, and $\hat{\mu}_{m,r}(x) = \sum_{j=1}^{m} \hat{p}_{j,r}(x)$ the estimated cumulative distribution function. Notice that, by simulation design, we can compute the true probabilities. I summarize these performance measures by averaging over the replications.

Table 4.1 displays the results obtained with $R = 400$ replications. The simulation shows that $MORF_A$ outperforms ORF uniformly in all the considered scenarios, as well as OL in *Design 2* and *Design 3*. Unsurprisingly, OL performs best in *Design 1*, where it correctly specifies the parametric model and the distributional assumption of the error term. However, its performance deteriorates when the model is misspecified, and the nonparametric methods ORF, $MORF_A$, and $MORF_H$ show superior results. The prediction performance of these

	Design 1				*Design 2*				*Design 3*			
	500	1,000	2,000	4,000	500	1,000	2,000	4,000	500	1,000	2,000	4,000
Panel 1: $\overline{MSE}$												
OL	0.018	0.009	0.004	0.002	0.075	0.069	0.065	0.064	0.172	0.166	0.163	0.162
ORF	0.134	0.121	0.110	0.100	0.068	0.058	0.050	0.043	0.121	0.108	0.096	0.086
$MORF_A$	0.127	0.114	0.103	0.094	0.057	0.047	0.040	0.034	0.107	0.095	0.085	0.077
$MORF_H$	0.168	0.149	0.136	0.126	0.079	0.067	0.058	0.050	0.138	0.125	0.112	0.102
Panel 2: $\overline{RPS}$												
OL	0.003	0.001	0.001	0.001	0.026	0.024	0.022	0.022	0.091	0.088	0.086	0.085
ORF	0.025	0.023	0.020	0.018	0.018	0.015	0.013	0.011	0.034	0.030	0.027	0.024
$MORF_A$	0.026	0.023	0.021	0.018	0.018	0.015	0.013	0.011	0.036	0.031	0.027	0.024
$MORF_H$	0.041	0.035	0.031	0.028	0.031	0.026	0.022	0.019	0.062	0.052	0.044	0.037

Table 4.1: Comparison with alternative estimators. The two panels report the average over the replications of MSE_r ($\overline{MSE}$) and RPS_r ($\overline{RPS}$).

methods improves in *Design 2* and *Design 3* compared to *Design 1*. This can be attributed to the general unsuitability of random forests for linear DGPs.

Although $\boldsymbol{MORF}_A$ and $\boldsymbol{ORF}$ achieve the same ranked probability scores, the former consistently exhibits a lower mean squared error across all considered scenarios. The advantage of $\boldsymbol{MORF}_A$ over $\boldsymbol{ORF}$ is relatively small in *Design 1*, where the MSE of $\boldsymbol{ORF}$ is between 5% and 7% larger than the MSE of $\boldsymbol{MORF}_A$. However, the superiority of $\boldsymbol{MORF}_A$ is more pronounced in *Design 2* (between 20–25%) and *Design 3* (between 12–13%).

Finally, we compare the prediction performance of $\boldsymbol{MORF}_A$ and $\boldsymbol{MORF}_H$ to quantify the cost of honesty resulting from using fewer observations to construct the forests. The results indicate that in terms of MSE, the cost ranges between 29%–47%, while in terms of RPS, the cost ranges between 50%–78%. Despite this cost, $\boldsymbol{MORF}_H$ still outperforms $\boldsymbol{OL}$ when the latter is misspecified.

4.2 Empirical Results

I utilize the same data sets considered by Janitza et al. (2016), Hornung (2020), and Lechner and Okasa (2019). These data sets differ in terms of the number of covariates and observations. Table 4.2 provides a summary of the data sets. For further details on the background of each data set, the reader is referred to Janitza et al. (2016).

To assess the prediction accuracy of each estimator, I employ a ten-fold cross-validation procedure. Specifically, I randomly divide each data set into ten folds $\mathcal{S}^1, \dots, \mathcal{S}^{10}$ with roughly equal sizes. For each fold $f = 1, \dots, 10$, I fit all the estimators using the observations

Data Sets

Data set	Sample Size	Outcome	Class range			N. Covariates
vlbw	218	Apgar score	1 (Life-threatening)	–	9 (Optimal)	10
mammography	412	Last mammography	1 (Never)	–	3 (Over a year)	5
support	798	Functional disability	1 (None)	–	5 (Fatal)	15
nhanes	1,914	Health status	1 (Excellent)	–	5 (Poor)	26
wines	4,893	Quality	1 (Moderate)	–	6 (High)	11

Table 4.2: Summary of data sets.

from all the other folds except for $\mathcal{S}^f$. Then, I calculate the same performance measures of Section 4.1 using the held-out $\mathcal{S}^f$:

$$MSE_f = \frac{1}{|\mathcal{S}^f|} \sum_{i \in \mathcal{S}^f} \sum_{m=1}^{M} \left[\mathbb{1} \left(Y_i = m \right) - \hat{p}_{m,f} \left(X_i \right) \right]^2 \tag{4.3}$$

$$RPS_f = \frac{1}{|\mathcal{S}^f|} \sum_{i \in \mathcal{S}^f} \frac{1}{M-1} \sum_{m=1}^{M} \left[\mathbb{1} \left(Y_i \leq m \right) - \hat{\mu}_{m,f} \left(X_i \right) \right]^2 \tag{4.4}$$

with $\hat{p}_{m,f}(\cdot)$ the estimated conditional probabilities using all the other folds except for $\mathcal{S}^f$, and $\hat{\mu}_{m,f}(x) = \sum_{j=1}^{m} \hat{p}_{j,f}(x)$ the estimated cumulative distribution function. Finally, I repeat this process ten times. This approach eliminates the dependence of the results on a particular training-validation sample split.

Figure 4.1 reports the results. It displays boxplots showing the median and interquartile range of the estimated mean squared error (upper panel) and ranked probability score

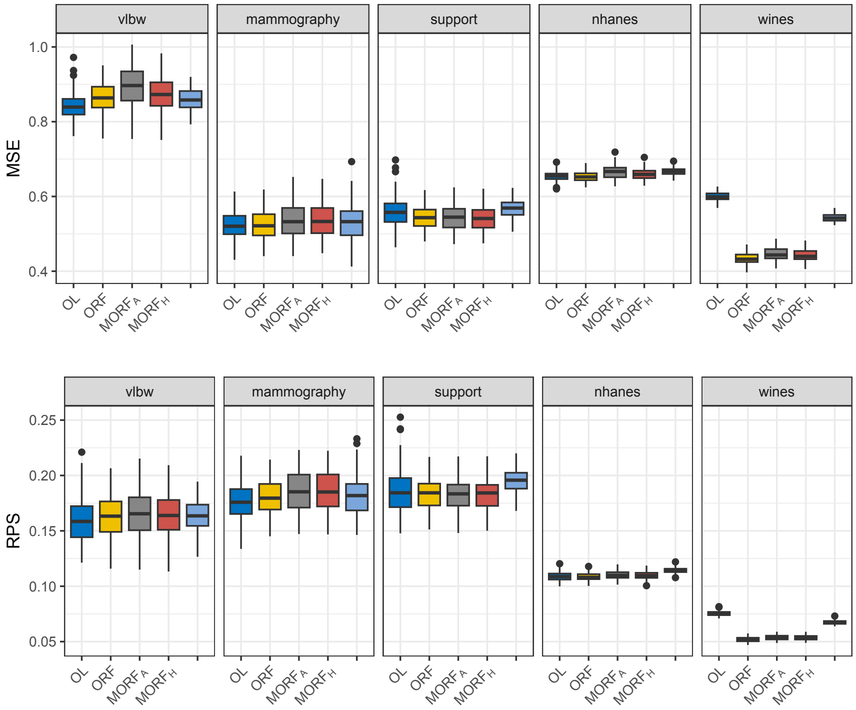

Figure 4.1: Prediction performance on real data sets. Each panel contains boxplots showing the median and interquartile range of the estimated mean squared error (upper panel) and ranked probability score (lower panel). The name of the data set is displayed at the top of each panel. Data sets are sorted according to their sample size.

(lower panel), together with their minima and maxima.[6] Overall, the results suggest that $\boldsymbol{MORF_A}$ performs competitively compared to the other estimators. $\boldsymbol{OL}$ exhibits slightly better performance in small data sets, where parametric methods are expected to perform well. However, as the sample size increases, its relative performance declines, and it ranks as the worst estimator in the largest sample size.

Consistent with the simulation results, the comparison of $\boldsymbol{MORF_A}$ and $\boldsymbol{ORF}$ in the real data sets reveals that both methods achieve similar ranked probability scores. However, $\boldsymbol{MORF_A}$ shows a slightly better prediction performance in terms of mean squared error in the *vlbw* and *nhanes* data sets, where the median MSE or $\boldsymbol{ORF}$ is between 1–3% larger than that of $\boldsymbol{MORF_A}$.

Finally, in contrast to the simulation results, the cost of honesty resulting from constructing the forests with fewer observations is relatively small and varies with sample size. Across the three largest data sets, the cost ranges between 1%–24% in terms of median MSE and between 4%–27% in terms of median RPS. In the two smallest data sets, $\boldsymbol{MORF_H}$ outperforms $\boldsymbol{MORF_A}$.

5 Conclusion

This paper proposes a novel estimator for ordered non-numeric outcomes, the *modified ordered random forest*. The estimator modifies a standard random forest splitting criterion to build a collection of forests, each estimating the conditional probability of a single class. Evidence from synthetic and real data sets shows that the proposed estimator features a superior prediction performance than alternative estimators.

Under an honesty condition, the proposed estimator inherits the asymptotic properties of random forests proven by Wager and Athey (2018), namely the consistency and asymptotic

[6] The cross-validation exercise yields a smaller sample size for estimating the expected MSE and RPS compared to the simulation results presented in the previous section. As a consequence, estimates of these targets can be more imprecise and influenced by outliers. I report the distribution of the estimated MSE and RPS using boxplots to provide a more robust assessment of the prediction performance of each estimator.

normality of its predictions. This allows valid inference about conditional probabilities to be made using conventional methods. Moreover, transforming the weights induced by each forest provides a methodology to obtain estimation and inference about the marginal effects of each covariate on the estimated probabilities.

References

Athey, S., & Imbens, G. W. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, *113*(27), 7353–7360.

Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests. *The Annals of Statistics*, *47*(2), 1148–1178.

Boes, S., & Winkelmann, R. (2006). Ordered response models. *Allgemeines Statistisches Archiv*, *90*(1), 167–181.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Wadsworth; Brooks.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32.

Cortez, P., Cerdeira, A., Almeida, F., Matos, T., & Reis, J. (2009). Modeling wine preferences by data mining from physicochemical properties. *Decision Support Systems*, *47*(4), 547–553.

Efron, B., & Hastie, T. (2016). *Computer age statistical inference: Algorithms, evidence, and data science*. Cambridge University Press.

Epstein, E. S. (1969). A scoring system for probability forecasts of ranked categories. *Journal of Applied Meteorology*, *8*(6), 985–987.

Frey, B. S., & Stutzer, A. (2002). What can economists learn from happiness research? *Journal of Economic literature*, *40*(2), 402–435.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of statistics*, 1189–1232.

Frijters, P., Haisken-DeNew, J. P., & Shields, M. A. (2004). Money does matter! evidence from increasing real income and life satisfaction in east germany following reunification. *American Economic Review*, *94*(3), 730–740.

Greene, W. H., & Hensher, D. A. (2010). *Modeling ordered choices: A primer*. Cambridge University Press.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer.

Hornung, R. (2020). Ordinal forests. *Journal of Classification*, *37*(1), 4–17.

Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical statistics*, *15*(3), 651–674.

Janitza, S., Tutz, G., & Boulesteix, A.-L. (2016). Random forest for ordinal responses: Prediction and variable selection. *Computational Statistics & Data Analysis*, *96*, 57–73.

Lechner, M., & Mareckova, J. (2022). Modified causal forest. *arXiv preprint arXiv:2209.03744*.

Lechner, M., & Okasa, G. (2019). Random forest estimation of the ordered choice model. *arXiv preprint arXiv:1907.02436*.

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

McCullagh, P., & Nelder, J. A. (2019). *Generalized linear models*. Routledge.

Murphy, A. H. (1970). The ranked probability score and the probability score: A comparison. *Monthly Weather Review*, *98*(12), 917–924.

Peracchi, F., & Rossetti, C. (2012). Heterogeneity in health responses and anchoring vignettes. *Empirical Economics*, *42*(2), 513–538.

Peracchi, F., & Rossetti, C. (2013). The heterogeneous thresholds ordered response model: Identification and inference. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *176*(3), 703–722.

Piccarreta, R. (2008). Classification trees for ordinal variables. *Computational Statistics*, *23*(3), 407–427.

Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, *113*(523), 1228–1242.

Wager, S., Hastie, T., & Efron, B. (2014). Confidence intervals for random forests: The jackknife and the infinitesimal jackknife. *The Journal of Machine Learning Research*, *15*(1), 1625–1651.

Wright, M. N., & Ziegler, A. (2017). Ranger: A fast implementation of random forests for high dimensional data in c++ and r. *Journal of Statistical Software*, *77*(1), 1–17.