STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **11**, September 2023, pp 865–875.
Published online in International Academic Press (www.IAPress.org)

Random forests in the zero to k inflated Power series populations

Hadi Saboori, Mahdi Doostparast*

Department of Statistics, Ferdowsi University of Mashhad, Mashhad, Iran

Abstract Tree–based algorithms are a class of useful, versatile, and popular tools in data mining and machine learning. Indeed, tree aggregation methods, such as random forests, are among the most powerful approaches to boost the performance of predictions. In this article, we apply tree–based methods to model and predict discrete data, using a highly flexible model. Inflation may occur in discrete data at some specific points such as zero, one or the others. We may even have inflation at two non-adjacent points or more. We use some recently introduced models for inflated data sets based on a common discrete family (the Power series models). The main idea of this article is to use zero to k ($k = 0, 1, \ldots$) inflated regression models based on the family of power series to fit decision regression trees and random forests. An important point of these models is that they can be used not only for inflated discrete data but also for non-inflated discrete data.

Keywords Random forest, Regression tree, zero to k inflated Power series model, Regression

AMS 2010 subject classifications 62-07, 62J12

DOI: 10.19139/soic-2310-5070-1773

1. Introduction

Count data has many applications in practice. Some examples are the number of warranty or insurance policy claims by a client, the number of unpaid credit installments, the number of accidents on a highway and the number of seizures for an epileptic. When the goal is to relate a set of covariates to the number of events for a sample of subjects, many parametric models are available, including the Poisson, negative binomial, Multinomial and Logarithmic series regression models (the Power series models regression). [12] provides tree using a nice treatment covering the modeling of count data.

We are more interested in a particular class of nonparametric models based on recursive partitioning also called tree-based models. Tree-based methods, or just trees, are valuable alternatives to parametric methods and are very popular among practitioners. Some of advantages are: (1) no need to specify a parametric form, (2) ability to automatically detect interactions, and (3) ease of interpretation and visualization. They were first developed to handle a categorical or a continuous outcome; See [5] for the early developments of the classification and regression tree (CART) paradigm and the earlier references. This paradigm builds a large tree by selecting the best split among all possible splits at each intermediate node. In order to avoid overfitting, a subtree is then selected by pruning the large tree using a cross validation mehtod. Within a similar framework as CART, Poisson regression trees can be fitted ([23]). In this case, the splitting criterion is on the basis of the likelihood ratio test to compare two Poisson distributions. However, a Bayes estimate of the rate is used in order to avoid an infinite value of the deviance which occurs when the maximum likelihood estimate of the rate is zero; See [23] for more details.

[9] proposed another method to build Poisson regression trees. It proceeds by fitting a log linear model in each node. The adjusted Anscombe residuals of the fitted model are then obtained. The Levene's two sample test is then

*Correspondence to: Mahdi Doostparast (Email: doustparast@um.ac.ir). Department of Statistics, Ferdowsi University of Mashhad, P.O. Box 91775-1159, Mashhad, Iran.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

applied to each covariate to compare the two groups formed by the positive and negative residuals. The selected covariate for the split is the one with the largest absolute statistic value. The split for the selected covariate is the average of the two group means along the covariate. Once a large tree is built, a pruning algorithm can be applied as in CART.

Extensions of Poisson regression trees into different directions have been proposed. [15] generalized the method proposed by [9] to the case of multivariate outcomes with models using the generalized estimating equations (GEEs). This method can be used to fit multivariate count data targets. In some applications, count data exhibit greater variability than what is expected from a Poisson model. To handle this, [10] extended both the [9] and The Generalized, Unbiased, Interaction detection and estimation (GUIDE) approach to the case of a Poisson outcomes with extra variation ([16] and [17]). In other applications, count data exhibit more zeros than what is expected from a Poisson model. [15], and [18], proposed a zero-inflated tree model with the CART approach.

In the last decade, the attention has shifted towards using trees as part of an ensemble learning algorithm. This is mainly due to the fact that combining many trees has often a better predictive capability than a single tree. Random forests ([6]) is such an ensembling method among the most popular ones. The good performance of random forests has been demonstrated in various empirical studies (e.g., [13] and [28]) and their theoretical properties have also been studied ([2], [3]). This is why they are now part of the standard practitioners' toolbox. For preliminary studies in random forests and ensemble methods, the articles introduced by [19], [21] and [24] are good starting points. [18] developed trees and random forests to predict the zero-inflated Poisson (ZIP) targets which considered nonhomogeneous Poisson processes. In fact, we combine two new approaches in fitting tree-base models.

In this paper, we fitted decision trees and random forests using regression models for responses where models belong to the zero to k inflated power series (ZKIPS) family of distributions. One is to use the power series family, which includes important models for discrete data, and the second is to generalize inflation points to $0, 1, \ldots, k$, $k = 0, 1, \ldots$. These two generalizations create two advantages. The former is the model's ability to fit count data that is inflated at any point. The later, due to the breadth of the power series family, as well as the structure of this model, is the ability to fit a wide range of discrete (even non-inflated) data.

The paper is organized as follows.In Section 2, we define the regression ZKIPS family. Section 3 describes the basic methodology of regression tree and random forest. The results from a simulation study are presented in Section 4. We evaluate the performance of proposed model with two real data sets, in Section 5. Concluding remarks and recommendations are stated in Section 6

2. Theoretical Backgrounds

First, we define the power series (PS) models. Suppose that the probability mass function (PMF) of the random variable Y is

$$p_\theta(y) = a(y)\theta^y / c(\theta), \tag{1}$$

where, $y \in \{0, 1, \ldots\}$, $a(y) > 0, \theta > 0$ and $c(\theta) = \sum_{y=0}^{\infty} a(y)\theta^y$. Then Y has the power series model and denoted by $Y \sim PS(\theta)$.

In standard model, assuming a linear predictor $X_i^\top \boldsymbol{\beta}$, where $X_i^\top$ is a matrix of regression covariates and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^\top$ is a parameter vector that describes the relation between the response (Y_i) and covariates $(X_i^\top)$, with $i = 1, 2, \ldots, n$, the link function $g(.)$ from Table 1, and θ_i the parameters of PS model, gives

$$\theta_i = g\left(X_i^\top \boldsymbol{\beta}\right) \Longleftrightarrow g^{-1}(\theta_i) = X_i^\top \boldsymbol{\beta}. \tag{2}$$

The family of PS models includes several important discrete models (Poisson, negative binomial, logarithmic, and multinomial models). On the other hand, by selecting the number of inflated points k, various sub-models are obtained. Table 2 summarizes some of the members of the ZKIPS models (r in the negative binomial model is the number of successes). We define the ZKIPS model as follows

Table 1. Form of the link function $g(.)$ for the family of the power series models

Distribution	$g(.)$
Multinomial	$\frac{\exp(X_i^\top \boldsymbol{\beta})}{1+\exp(X_i^\top \boldsymbol{\beta})}$
Negative binomial	$\frac{\exp(X_i^\top \boldsymbol{\beta})}{1+\exp(X_i^\top \boldsymbol{\beta})}$
Poisson	$\exp\left(X_i^\top \boldsymbol{\beta}\right)$
Logarithmic series	$\frac{\exp(X_i^\top \boldsymbol{\beta})}{1+\exp(X_i^\top \boldsymbol{\beta})}$

Table 2. The family of the power series models with form (1)

Distribution	θ_i	$c(\boldsymbol{\theta})$	$a(y_1, \cdots, y_m)$
Binomial	$\theta_i = \frac{p_i}{1-p_i}$	$(1-p_i)^n$	$\frac{n!}{y_i!(n-y_i)!}$
Negative binomial	$\theta_i = p_i$	$(1-p_i)^{-r}$	$\frac{\Gamma(y_i+r)}{y_i!\Gamma(r)}$
Poisson	$\theta_i = \lambda_i$	$\exp(\theta_i)$	$1/y_i!$
Logarithmic	$\theta_i = p_i$	$-\log(1-p_i)$	$\frac{(y_i-1)!}{y_i!}$

$$Pr(Y_i = y_i | X_i^\top, \xi_i) = \begin{cases} w_0 + \eta(\mathbf{w}) p_{\theta_i}(0), & y_i = 0 \\ w_1 + \eta(\mathbf{w}) p_{\theta_i}(1), & y_i = 1 \\ \vdots & \vdots \\ w_k + \eta(\mathbf{w}) p_{\theta_i}(k), & y_i = k \\ \eta(\mathbf{w}) p_{\theta_i}(y_i), & y_i > k, \end{cases} \tag{3}$$

where $0 \leq w_i, i = 0, .., k$, $\eta(\mathbf{w}) = 1 - \sum_{i=0}^{k} w_i$, $0 \leq \eta(\mathbf{w}) \leq 1$, $\mathbf{w} = (w_0, \cdots, w_k)$, and $p_\theta(.)$ is the PS model PMF (1). Each $Y_i, i = 1, \cdots, n$ follows the ZKIPS distribution; i.e. $Y_i \sim$ ZKIPS$(\mathbf{w}, \theta_i)$, with PMF, (3).
Some special cases are:

- $k = 0 \rightarrow$ zero-inflated Power series model.
- $k = 1 \rightarrow$ zero-one inflated Power series model.
- $k = 2 \rightarrow$ zero-one-Two inflated Power series model.
- $k = 3, w_1 = w_2 = 0 \rightarrow$ zero-three inflated Power series model.

Suppose that we have a vector $\mathbf{y} = (y_1, \ldots, y_n)$ of n independent responses from ZKIPS models. Associated with each y_i, we have a vector of p covariates $\mathbf{X}_i^\top = (x_{1i}, \ldots, x_{pi})^\top$ of features have been also observed. From (3), the likelihood function (LF) of the available data

$$L(\mathbf{w}, \boldsymbol{\theta}; \mathbf{y}) = \prod_{j=1}^{n} \left\{ \prod_{i=0}^{k} \left(\left[w_i + \eta(\mathbf{w}) P_{\theta_j}(y_j) \right]^{I_i(y_j)} \right) \times \left(\eta(\mathbf{w}) P_{\theta_j}(y_j) \right)^{I(y_j > k)} \right\}, \tag{4}$$

where $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_n)$ and

$$I_i(a) = \begin{cases} 1, & a = i \\ 0, & a \neq i, \end{cases} \quad i = 0, \ldots, k, \tag{5}$$

and $I(a > k) = 1 - \sum_{i=0}^{k} I_i(a)$. The log likelihood function (LLF) of the LF (4) is

$$\ell(\mathbf{w}, \boldsymbol{\theta}; \mathbf{y}) = \log\left[L(\mathbf{w}, \boldsymbol{\theta}; \mathbf{y})\right] = \sum_{j=1}^{n} \left\{ \sum_{i=0}^{k} I_i(y_j) \left(w_i + \eta(\mathbf{w}) P_{\theta_j}(y_j)\right) + I(y_j > k) \log\left(\eta(\mathbf{w}) P_{\theta_j}(y_j)\right) \right\}. \tag{6}$$

By Equation (6), the maximum likelihood estimates (MLE) are derived by solving the following equations with respect to the parameters:

$$\frac{\partial \ell(\mathbf{w}, \boldsymbol{\theta}; \mathbf{y})}{\partial w_s} = \sum_{j=1}^{n} \left\{ I_s(y_j) \left(1 - P_{\theta_j}(y_j)\right) + \frac{I(y_j > k)}{\eta(\mathbf{w})} \right\} = 0, \quad s = 0, \ldots, k, \tag{7}$$

$$\frac{\partial \ell(\mathbf{w}, \boldsymbol{\theta}; \mathbf{y})}{\partial \theta_l} = \eta(\mathbf{w}) \sum_{i=0}^{k} I_i(y_l) \frac{\partial P_{\theta_l}(y_l)}{\partial \theta_l} + I(y_l > k) \frac{\partial P_{\theta_l}(y_l)/\partial \theta_l}{P_{\theta_l}(y_l)} = 0, \quad l = 1, \ldots, n, \tag{8}$$

and with suppose $\theta_1 = \ldots = \theta_n = \theta$, (models are identical), we have

$$\frac{\partial \ell(\mathbf{w}, \boldsymbol{\theta}; \mathbf{y})}{\partial \theta} = \sum_{j=1}^{n} \left\{ \eta(\mathbf{w}) \sum_{i=0}^{k} I_i(y_l) \frac{\partial P_{\theta_l}(y_j)}{\partial \theta} + I(y_j > k) \frac{\partial P_{\theta}(y_j)/\partial \theta}{P_{\theta}(y_j)} \right\} = 0, \tag{9}$$

where, the form of $P_\theta(y)$ comes from (1) and Table 2.
To connect the parameters with the covariates, we follow the standard generalized linear model (GLM) framework; see [1] for further reading.

3. Decision trees and forest algorithms

Tree-based methods partition the covariate space by splitting it recursively with rules based on covariates. The basic ingredient for building a tree is the splitting criterion, which is a dependent-problem. For example, the least squares splitting criterion is the usual one when the response is continuous. Suppose we are at a given node t and we want to split it into two children nodes, t_L (left node) and t_R (right node). The best split is chosen among all possible binary splits obtained from a covariate, say X_i. If X_i is continuous, then $\{X_i < s\}$ and $\{X_i \geq s\}$. If X_i is categorical, the possible splits take the form $X_i \in \{c_1, \ldots, c_l\}$ where $\{c_1, \ldots\ldots, c_l\}$ is a subset of the possible values of X_i. The best split is the one maximizing a given criterion. If a single tree is required, then the usual procedure builds a large tree and then uses a pruning algorithm to avoid overfitting. However, it is now well-established that using an ensemble of trees is generally preferable to a single tree. One of the most popular ensemble method is random forests, introduced by [6]. Here, we describe the generic random forest algorithm that will be used in this paper:

Algorithm:

For $b = 1, \ldots, B$ bootstrap samples $\{\mathbf{y}_b, \mathbf{X}_b\}$:

Step 1. From the training dataset, draw bootstrap samples of size N_{train}.

Step 2. With the bootstrapped data, grow a tree T_b with the maximum likelihood splitting criterion, by recursively repeating the following steps for each terminal node of the tree, until the minimum node size is

reached.

1) Randomly draw $mtry$ out of the m independent variables (IVs). $mtry$ is a user-specified parameter.

2) Pick the best independent variable among the m-try IVs.

3) Split the node into two child nodes. The split ends when a stopping criterion is reached, for instance, when a node has less than a predetermined number of observations. No pruning is performed.

Step 3. Output the ensemble of trees $\{T_b\}_1^B$.

The predicted value of testing set ($\hat{y}_i$) individuals with input $\mathbf{X}_i$ is calculated as $\hat{y}_i = \frac{1}{B}\sum_{b=1}^{B} T_b(X_i)$. Readers are referred to [6] and [25] for details on the theory of RF. Tree hyper-parameters, including the number of trees (ntree), number of independent variables (features) sampled in each iteration (mtry), and number of samples in the final nodes (nodesize) must be defined by the user.

3.1. Maximum likelihood splitting criterion

A simple method for deriving a splitting criterion is to use the log-likelihood of an adequate two-node model; see [22] and [4]. The main idea of this article is use of the LLF of regression ZKIPS model for the splitting criterion function. Basically, the best split at a given node is the one that maximizes the observed LLF, i.e. the one evaluated at the MLE, among all allowable splits. Moreover, if the parameters are estimated separately in the two children nodes, then the best split is the one that maximizes

$$\widehat{LL}(\textit{left node}) + \widehat{LL}(\textit{right node}), \tag{10}$$

where $\widehat{LL}(\textit{left node})$ and $\widehat{LL}(\textit{right node})$ are the observed log-likelihood in the left and right nodes, respectively. [15] proposed a decision tree method for zero-inflated count data based on the CART paradigm. They call it a ZIP tree. They basically fit the ZIP distribution separately in the two children nodes, and use (10) as the splitting criterion. We use a similar criterion, but based on the ZKIPS model. Similar to [15], we consider the homogeneity measure using LLF (6).

3.2. Tree pruning

The aforementioned splitting method is applied recursively until one of the stopping criteria is reached. At each node, common stopping criterion, like all covariates are the same or all the response values are the same, is adopted. Moreover, node t_g is considered terminal, if the number of cases in t is less than 5% of the total observation in the training data set. Afterwards, the cost-complexity pruning method ([5]) is adopted to find the right-sized tree. At each node, the node deviance is used as the error measure in the pruning algorithm ([5], Chapter 10). The tree is then pruned by the cost-complexity method with ten-fold cross-validation. Node deviance has been used to prune trees in [8] among others. In simulation studies and real data analyses for fitting regression tree, we use the maximum likelihood splitting criterion

4. A simulation study

In this section, we conduct to a simulation study examine the performance of family ZKIPS. As an example, we consider zero-one inflated Negative binomial (Z1INB) mode. This simulation can be done similarly for the rest of the family. In addition, this simulation is used only to implement a regression tree, which can be generalized to random forests.

Table 3. AAE and ARE for some selected response models

Distribution	AAE	ARE
Z1INB	1.1921	0.3491
ZINB	1.1992	0.3496
NB	1.9608	0.3508
Z1IP	1.2027	0.3532
Z1IB	1.1933	0.3556

4.1. Simulation details

To perform this simulation, we first considered 5 covariates independently of the possible distributions of $X_1 \sim$ Normal$(0,1)$, $X_2 \sim$ DUniform$\{-1,1\}$, $X_3 \sim$ Poisson(1), $X_4 \sim$ Beta$(5,2)$, and $X_5 \sim$ Uniform$(-1,1)$, respectively. Where, Normal, DUniform, Poisson, Beta, and Uniform respectively are Normal, Discrete uniform, Poisson, Beta, and Continuous uniform distributions.
In each step, and using (2), we produce a sample of size 100 from the Z1INB ($r = 5$) model. Then, we fitted the regression tree based on distributions Z1INB, zero inflated Negative binomial (ZINB), Negative binomial (NB), zero-one inflated Poisson (Z1IP) and zero-one inflated Binomial (Z1IB) models for this data set. We repeated this step 500 times. Given that the data are generated from Z1INB, it seems that the regression tree implemented using this model has the best performance, which is confirmed by the results of Table 3. The comparison between the results of regression tree fitting was performed using two criteria average absolute error (AAE) and average relative error (ARE) defined by

$$\begin{aligned} AAE &= 1/n \sum_{i=1}^{n} |y_i - \hat{y}_i|, \\ ARE &= 1/n \sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{y_i + 1}, \end{aligned}$$

which $\hat{y}_i$ is the model prediction value for the observed response variable $i = 1, \cdots, n$, and n is the size of data set. According to Table 3, the values of two criteria AAE and ARE for the fitted regression tree based on model Z1INB have the lowest values.

5. Applications

We used two real data sets to evaluate performance of the ZKIPS models. In the following, the aim is to show how to choose the best model for a real data analysis from the inflated power series family of models.

5.1. Solder data set

This data was first used by [11]. This data is available in the R package rpart. It contains 720 observations which are the result of an experiment on wave-soldering of electronic components in a printed circuit board. The response variable is the number of solder skips. The covariates are: Mask, type and thickness of the material for the solder mask (5 levels), Opening, thickness of clearance around a mounting pad (3 levels), Pad type, geometry and size of the mounting pad (10 levels), Panel, panel position on a board (3 levels) and Solder, thickness of solder (2 levels). We fitted three important models of the power series family to this data (the logarithmic series model was omitted because this model is not defined at the zero point because the data contains zero-response observations). In the Binomial model, parameter $n = 50$ and in the Negative binomial model, parameter $r = 1$ (which results in the best

Table 4. The family of the power series models for the Solder data set.

Distribution		Without inflation	$k=0$	$k=1$	$k=2$	$k=3$
Binomial	Deviance	9037.806	4816.001	4208.780	3765.721	3571.482
$(n=50)$	AIC	18077.611	9636.002	8423.561	7539.441	7152.964
Negative binomial	Deviance	6289.775	2423.172	2406.438	2394.451	2394.297
$(r=1)$	AIC	12581.550	4850.344	4818.877	4796.901	4798.594
Poisson	Deviance	5445.399	4112.004	3592.179	3228.262	3078.748
	AIC	10892.797	8228.009	7190.357	6464.525	6167.496

Table 5. The family of the power series models for the Solder data set, regression decision tree.

Distribution		Without inflation	$k=0$	$k=1$	$k=2$	$k=3$
Binomial	SSE	33.012	49.778	33.051	33.084	33.084
$(n=50)$	Deviance	8083.694	7765.285	11319.760	6048.475	7484.670
Negative binomial	SSE	33.012	33.012	33.012	33.012	33.012
$(r=1)$	Deviance	5104.882	5582.121	5582.644	5521.462	5524.705
Poisson	SSE	33.012	33.012	33.051	33.084	33.084
	Deviance	3678.175	8695.353	10227.88	6782.28	6935.557

performance of the model) are considered. The results are presented in Tables 4-6. First, in Table 4, we compare the three models of the Binomial, Negative binomial, and Poisson in non-inflated and inflated cases (in $k=0,1,2,3$). These comparisons, which were performed using two criteria Deviance and the Akaike information criterion (AIC), indicate that each model has approximately the $k=2$ best fit. In total, the zero to two inflated negative binomial (Z2INB) model offers the best fit. Of course, in the Negative binomial model, the criterion Deviance in $k=3$ is slightly smaller than $k=2$, but the criterion of AIC in $k=2$ is smaller. Finally, Table 4 offers the Model 2 as a suitable model for fitting this data. Finally, Table 4 offers model Z2INB as the best model in the family of power series to fit this data set.

In Tables 5 and 6, we fitted a regression decision tree, and random forests to the data. The comparison in these two tables is done using three criteria: the sum of squared error (SSE), Akaike information criterion (AIC), and Deviance.

$$\begin{aligned} SSE &= \sum_{i=1}^{n}(y_i-\hat{y}_i)^2, \\ AIC &= 2s-2LLF, \\ Deviance &= -2LLF, \end{aligned}$$

where y_i, the response, $\hat{y}_i$, estimation of response by model, n, the number of observation and s is number of estimated parameters in the model.

To fit the decision tree, we considered the minimum data per node to be 10 observations (besides, the fitted tree was not so complex that it needed pruning). Again, the results of Table 4 are confirmed and in Table 5, the least SSE and Deviance belong to the Z2INB model.

To fit random forests, we used the decision trees used in Tables 5 and 6 covariates for each fit (that are randomly selected from among the covariates each time). The results of random forests fitting show that the fitting has improved compared to Table 5, but the overall results are almost the same as before. Finally, the best model to fit this data is the Z2INB model. the decision tree Z2INB fits for this data and variable importance ([14]) for the obtained decision tree and random forests are shown in Figure 1.

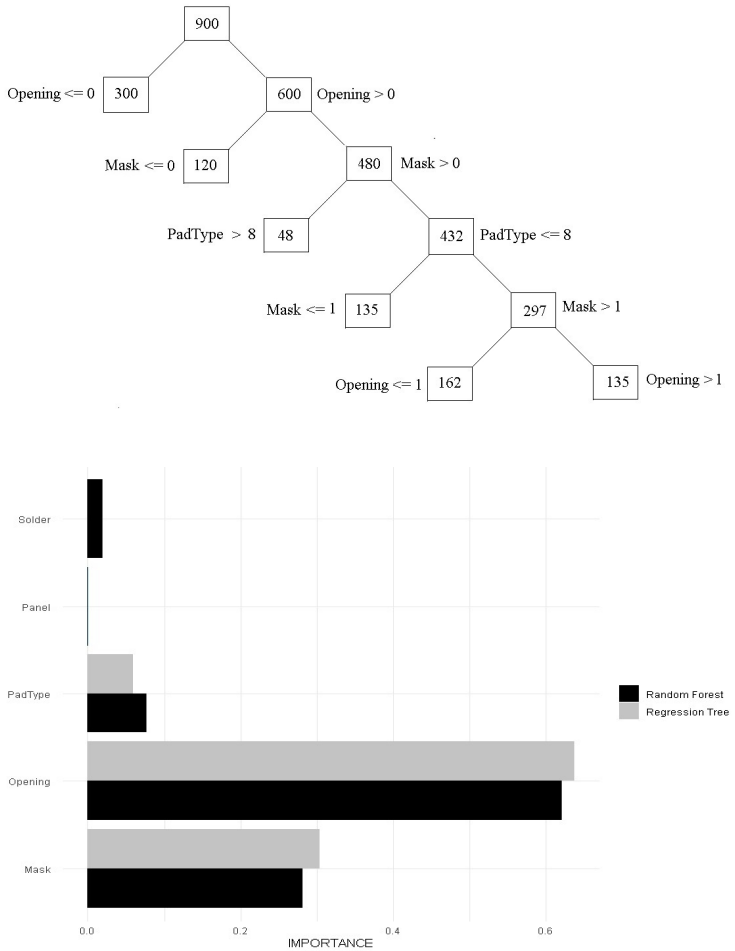

Figure 1. (a) Final fitted tree for Solder data set and (b) variable importance for the fitted decision tree and random forests

Table 6. The family of the power series models for the Solder data set, random forest.

Distribution		Without inflation	$k=0$	$k=1$	$k=2$	$k=3$
Binomial	SSE	33.012	49.778	33.051	33.653	32.528
($n=50$)	Deviance	6965.112	7090.248	10560.023	6048.278	7636.723
Negative binomial	SSE	33.888	33.907	32.938	30.696	31.279
($r=1$)	Deviance	5052.24	5568.667	5535.968	5450.380	54.68.713
Poisson	SSE	32.404	33.911	35.057	32.423	36.681
	Deviance	7353.147	7906.080	9303.755	6649.066	5929.916

5.2. Health care data set

There are numerous studies of the impact of insurance on health care use, measured by the number of services. For example, [26] and [27] studied the impact of health reform on doctor visits in Germany, using data coming from the German Socioeconomic Panel. [7] analyzed count data from the Rand Health Insurance Experiment. The response is the number of illnesses in the past two weeks for a single-adult ($n = 5,190$), from the Australian health

survey 1977–78, ([7]). The covariates are hscore, general health questionnaire score using Goldberg's method, chcond1, equals 1 if chronic condition not limiting activity, chcond2, equals1 if chronic condition limiting activity , medicine, number of medicines taken, prescribe, number of prescribe.
In the Binomial model, parameter $n = 5$ and in the Negative binomial model, parameter $r = 4$ (which results in the best performance of the model) are considered. With a similar analysis to the previous example, Table 7 considers the zero-5 inflated binomial (Z5IB) model appropriate for this data set. The results of decision tree fitting and random forests are presented in Tables 8 and 9, respectively. The indicators in these two tables almost confirm the results of Table 7. So it seems that to fit the decision tree as well as random forests on this data set, the best model of the power series family is the Z5IB model. the decision tree Z5IB fits for this data and variable importance for the obtained decision tree and random forests are shown in Figure 2.

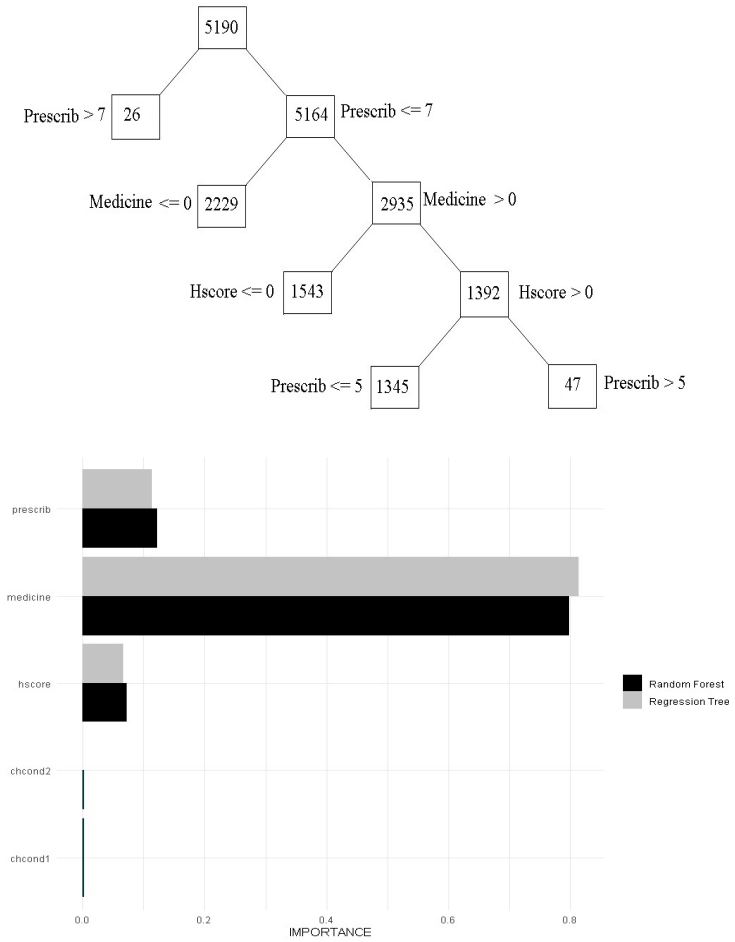

Figure 2. (a) Final fitted tree for the Health care data set and (b) Variable importance for fitted decision tree and random forests

6. Conclusions and remarks

In this paper, we tried to demonstrate the efficiency of the family of Power series in fitting regression trees. For this purpose, we first defined the family of inflated power series. We demonstrate that the most important feature of

Table 7. The family of the power series models for the Health care data set.

Distribution		Without inflation	$k=4$	$k=5$	$k=6$
Binomial	Deviance	9037.806	8133.159	8133.159	8133.165
$(n=5)$	AIC	18077.610	16278.320	16280.320	16282.330
Negative binomial	Deviance	8269.373	8255.729	8133.197	8139.078
$(r=4)$	AIC	16540.750	16523.460	16280.390	16294.160
Poisson	Deviance	8390.942	8235.029	8133.159	8133.160
	AIC	16783.880	16482.060	16280.320	16282.320

Table 8. The family of the power series models for the Health care data set, regression decision tree.

Distribution		Without inflation	$k=4$	$k=5$	$k=6$
Binomial	SSE	1.800	1.521	1.561	1.561
$(n=5)$	Deviance	11202.620	14665.880	13793.220	13796.500
Negative binomial	SSE	1.595	1.561	1.561	1.553
$(r=)$	Deviance	13861.880	13681.360	13757.450	13363.070
Poisson	SSE	1.561	1.561	1.567	1.567
	Deviance	12325.120	13494.460	15378.900	13586.950

Table 9. The family of the power series models for Health care data set, random forest.

Distribution		Without inflation	$k=4$	$k=5$	$k=6$
Binomial	SSE	1.686	1.470	1.501	1.456
$(n=5)$	Deviance	11468.580	14055.840	14658.160	14329.610
Negative binomial	SSE	1.566	1.477	1.358	1.472
$(r=4)$	Deviance	13307.330	14066.570	14542.120	14328.040
Poisson	SSE	1.499	1.471	1.507	1.490
	Deviance	12962.050	13819.110	15378.90	14667.750

this family is the ability to capture various kinds of inflated and even non-inflated data set. Therefore, considering the wide range of applications and flexibility of this family, it seems obvious that it is an appropriate option for use in fitting regression decision trees and, consequently, random forests. This distribution can be also used in boosting algorithm. The multivariate generalization of this distribution ([20]) can be also used to fit decision trees and random forests on multivariate data.

REFERENCES

1. A. Agresti, *Analysis of ordinal categorical data*, 3nd edn. Wiley, Hoboken, 2013.
2. G. Biau, L. Devroye, and G. Lugosi, *Consistency of Random Forests and Other Averaging Classifiers*, Journal of Machine Learning Research vol. 9, pp. 2015–2033, 2008.
3. G. Biau, and L. Devroye, *On the Layered Nearest Neighbour Estimate, the Bagged Nearest Neighbour Estimate and the Random Forest Method in Regression and Classification*, Journal of Multivariate Analysis vol. 101, pp. 2499–2518, 2010.
4. I. Bou-Hamad, D. Larocque, H. Ben-Hameur, L. C. Masse, F. Vitaro, and R. E. Tremblay, *Discrete-Time Survival Trees*, Canadian Journal of Statistics, vol. 37, pp. 17–32, 2009.
5. L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees. Belmont*, CA: Wadsworth International Group, 1984.
6. L. Breiman, *Random Forests*, Machine Learning, vol. 45, pp. 5–32, 2001.
7. A. C. Cameron, and P. K. Trivedi, *Regression analysis of count data*, Cambridge university press, vol. 53, 2013.

8. K. Y. Chan, and W. Y. Loh, *LOTUS: an algorithm for building accurate and comprehensible logistic regression trees*, Journal Comput Graph Stat. vol. 13, no. 4, pp. 826–852, 2004.
9. P. Chaudhuri, W. D. Lo, W. Y. Loh, and C. C. Yang, *Generalized Regression Trees*, Statistica Sinica, vol. 5, pp. 641–666., 1995.
10. Y. Choi, H. Ahn, and J. J. Chen, *Regression Trees for Analysis of Count Data with Extra Poisson Variation*, Computational Statistics and Data Analysis, vol. 49, pp. 893–915, 2005.
11. R. B. Comizzoli, J. M. Landwehr, and J. D. Sinclair, *Robust materials and processes: key to reliability*, AT and T Tech J, vol. 69, no. 6, pp. 113–128, 1990.
12. R. J. Cook, and J. F. Lawless, *The Statistical Analysis of Recurrent Events*, Springer. New York, 2007.
13. Y. Huang *Predicting Home Value in California, United States via Machine Learning Modeling*, Statistics, Optimization and Information Computing, vol. 7(1), pp. 66 – 74, 2019.
14. H. Ishwaran *Variable importance in binary regression trees and forests. Electron*, Journal Statist, vol. 1, pp. 519 – 537, 2007.
15. S. K. Lee, and S. Jin, *Decision Tree Approaches for Zero–inflated Count Data*, Journal of Applied Statistics, vol. 33, pp. 853-865, 2006.
16. W. Y. Loh, *Regression Trees With Unbiased Variable Selection and InteractionDetection*, Statistica Sinica, vol. 12, pp. 361-386, 2002.
17. W. Y. Loh, *Classification and Regression Trees*, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 1, pp. 14–23, 2011.
18. W. Mathlouthi, M. Fredette, and D. Larocque, *Regression trees and forests for non-homogeneous Poisson processes*, Statistics and Probability Letters, vol. 96, pp. 204–211, 2015.
19. L. Rokach, *Taxonomy for Characterizing Ensemble Methods in Classi cation Tasks: A Review and Annotated Bibliography*, Computational Statistics and Data Analysis, vol. 53, pp. 4046–4072, 2008.
20. H. Saboori, and M. Doostparast, *Flexible multivariate zero to k inflated power series regression model with applications*, Stat, vol. 11, no. 1, pp. e473, 2022.
21. D. S. Siroky, *Navigating Random Forests and Related Advances in Algorithmic Modeling*, Statistics Surveys, vol.3, pp. 147–163, 2009.
22. X. Su, M. Wang, and J. Fan, *Maximum Likelihood Regression Trees*, Journal of Computational and Graphical Statistics, vol. 13, pp. 586–598, 2004.
23. T. M. Therneau, B. Atkinson, and B. Ripley, *rpart: Recursive Partitioning and Regression Trees*, R package version 4.1-8, 2014. Verikas, A., Gelzinis, A. and Bacauskiene, M. (2011). . .
24. A. Verikas, A. Gelzinis, and M. Bacauskiene, *Mining Data With Random Forests: A Survey and Results of New Tests*, Pattern Recognition, vol. 44, pp. 330–349, 2011.
25. P. Waldmann, *Genome-wide prediction using Bayesian additive regression trees*, Genetics Selection Evolution, 48, pp. 1–12, 2016.
26. R. Winkelmann, *Health Care Reform and the Number of Doctor Visits – An Econometric Analysis*, Journal of Applied Econometrics, vol. 19, pp. 455–472, 2004.
27. R. Winkelmann, *Reforming Health Care: Evidence from Quantile Regressions for Counts*, Journal of Health Economics, vol. 25, pp. 131–145, 2006.
28. S. Zhong, and D. Hitchcock, *S&P 500 Stock Price Prediction Using Technical, Fundamental and Text Data*, Statistics, Optimization and Information Computing, vol. 9(4), pp. 769–788, 2021.