IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received January 11, 2022, accepted January 19, 2022, date of publication January 26, 2022, date of current version February 4, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3146406

Optimal Causal Decision Trees Ensemble for Improved Prediction and Causal Inference

NEELAM YOUNAS[1], AMJAD ALI[2], HAFSA HINA[1], MUHAMMAD HAMRAZ[2], ZARDAD KHAN[2], AND SAEED ALDAHMANI[3]

[1]Department of Econometrics, Pakistan Institute of Development Economics, Islamabad 44000, Pakistan
[2]Department of Statistics, Abdul Wali Khan University Mardan, Mardan 23200, Pakistan
[3]Department of Analytics in the Digital Era, United Arab Emirates University, Al Ain, UAE

Corresponding authors: Zardad Khan (zardadkhan@awkum.edu.pk) and Amjad Ali (amjad.ali@awkum.edu.pk)

ABSTRACT Ensemble methods can be used to identify causal relationships in data for a better understanding and taking the right decision in processes that involve high risk. This paper explores the idea of a causal decision tree forest and proposes a regularized ensemble method by integrating optimal causal trees for improved prediction accuracy while not compromising on accurately estimating heterogeneous treatment effects. The proposed method is based on selecting a subset of the most accurate causal trees from a sufficiently large pool based on their out-of-sample error estimates. The selected trees are integrated to form an ensemble that is used for estimating heterogeneous treatment effect and predicting unseen data. The proposed method is applied on Pakistan's income function consisting of 27964 observations on wages of workers age 10 and above as an example dataset. The paper gives a detailed simulation study where datasets are generated under 5 different designs. The proposed method is assessed against ordinary least square (OLS), least absolute shrinkage and selection operator (LASSO), Ridge, Causal Tree and the standard decision trees forest (i.e. the causal forest) via mean square error (*MSE*), root mean square error (*RMSE*), mean absolute deviation (*MAD*) and Pearson correlation (*r*) as performance metrics. The analyses given in the paper reveal that the proposed method can be used effectively for estimating heterogeneous treatment effects and achieves better prediction performance and as compared to the rest of the methods given in the paper.

INDEX TERMS Causal inference, causal decision tree, random forest, causal random forest, ensemble learning, heterogeneous treatment effect.

I. INTRODUCTION

The identification of the causal relationships in the data is key to provide a better understanding and the knowledge for taking an accurate decision in processes with risk. Such types of relationships are usually established with the help of experiments which are effective but, at the same time, costly and difficult to conduct [1]. Observational studies can also be used to find the causal relationships in the data [2], which are tested by taking a sample from historical data or by observing the characteristic of interest over a period of time, thereby making the observational study time-consuming.

Machine learning is generally used for accurately predicting unknown data based on learning from known data. However, sometimes, the purpose of using machine learning methods could potentially exceed prediction, such as representing and discovering causal relationships in data and estimating heterogeneous causal effects. This kind of application provides a compact and precise graphical representation of the causal relationships between a set of predictor attributes and an outcome attribute. Various machine learning methods are in use for obtaining the desired results. Typical examples include classification and regression trees, k-nearest neighbours models, support vector machines, etc.

The methods for the identification of causal relationship in the data should be capable of identifying the causal effect without any prior knowledge. Moreover, these methods should be capable of dealing high dimensional data sets efficiently. The methods of classification such as decision trees [3] have the ability to identify the causal relationships in the data by using a supervised learning approach where the response variable is known or fixed. Such type of methods are commonly used in medical and social data analyses, for example. However, these methods are not

The associate editor coordinating the review of this manuscript and approving it for publication was Ehab Elsayed Elattar.

specifically designed for identifying the causal relationships in the data, and hence can provide incorrect estimates of causal relationships.

Classification methods, like random forest, are fast and could find causal signals in the data effectively. These methods, with the provision of scalability and automation, are important for exploring causal relationships in large datasets. To this end, researchers have proposed several machine learning methods for exploring causal relationships. Although these methods serve the purpose of finding causal signals, they fail to achieve higher accuracy. Therefore, the aim of this paper is to achieve both causal exploration and high prediction accuracy by proposing a causal decision trees ensemble. This will help economists in predicting and answering various causal questions for policy implementation in a machine learning framework.

Various authors have suggested that combining weak models leads to efficient ensembles. Moreover, combining the outputs of multiple classifiers also reduces generalisation error. Ensemble methods have high efficacy in that the different models involved have different inductive biases, where such diversity reduces variance-error while not increasing the bias error [4]–[6]. As the number of trees in a random forest is often very large, there has been a significant work conducted on the problem of minimising this number to reduce computational cost without decreasing prediction accuracy [7]–[10]. The overall prediction error of a random forest is highly associated with the strength of individual trees and their diversity in the forest. This idea is backed by Breiman's [11] upper bound for the overall prediction error of random forest given by $\overline{Err} \leq \overline{\rho} err_t$, where $t = 1, 2, \ldots, B$, and B denotes the number of trees in the forest, $\overline{Err}$ is the overall prediction error of the forest and $\overline{\rho}$ represents weighted correlation between residuals from two independent trees, i.e. mean (expected value) of their correlation over entire ensemble, and err_t is the prediction error of some t^{th} tree in the forest.

Generally, a random forest is based on a large number of base trees, and researchers have always tried to minimise this number in order to gradually shrink the cost of computation without negatively affecting the prediction accuracy. Overall, the prediction error of a random forest is strongly connected with the accuracy of individual trees and their diversity in the forest. The proposed method selects a subset of the best causal trees in terms of their individual strength, i.e. accuracy, from a large ensemble grown by the causal random forest. The selected trees are combined in an ensemble for predicting unknown data and estimating heterogeneous treatment effects in the data. The proposed method is applied on an example dataset from the labour force survey (LFS) of Pakistan for causal effects exploration. The paper also gives a detailed simulation study of the proposed method in comparison with causal decision tree, causal random forest, ordinary least square (OLS) linear regression, least absolute and shrinkage and selection operator (LASSO) and ridge regression for further assessment. For judging the efficacy of the newly developed method, conditional average treatment effect (*CATE*), average treatment effect (*ATE*), mean square error (*MSE*), root mean square error (*RMSE*), mean absolute error (*MAE*) and Pearson correlation coefficient (r) are used as performance measures.

The remainder of this paper is organised as follows. Section II provides a summary of the related work done in the literature; Section III presents a detailed description of the method proposed in this paper; Section IV gives the analyses conducted in this paper based on the simulated and real datasets; and Section V concludes the findings.

II. RELATED WORK

Extensive research has been done in the literature for estimating the parameters of interest, like heterogeneous treatment effects. Some well known methods consist of local maximum likelihood and local generalized method of moments such as [12]–[17]. Some applications of these techniques in the field of economics includes multinomial choice models in a longitudinal data type i.e. [18] and instrumental variables regression i.e. [19]. To estimate the parameters at a particular value of covariates, the core idea is to use kernel weighting function in order to place more weight on nearby observations in the covariate space. The main problem in such types of techniques is that, if the feature space is high dimensional, then the performance of these methods can be suffered from the problem known as ''curse of dimensionality'' [20].

Authors in [21] replaced kernel weighting with the forest-based weights i.e. weights that are obtained from the trees fraction that contain observation in the same leaf as the response value of covariate vector. Study in [11] proposed a random forest algorithm for non-parametric classification and regression building on insights from the ensemble learning literature [22]–[25]. Random forest as a type of adaptive nearest neighbor prediction is closely built in the studies given in [26] and [27], which is forest-based method for quantile regression and survival analysis. Authors in [28]–[34], have used gradient based test statistics to identify the change points in likelihood models.

According to [35], numerous data mining techniques such as classification, k-nearest neighbors and sequential pattern mining are sequentially applied to identify the similarity in decision trees. Also authors in [36] used unified Granger causality analysis (uGCA) framework for sequential medical imaging. The study in [37] proposed hierarchical probabilistic graphical model to simultaneously handle classification of multi-sensor and multi-resolution remote sensing of the same scene. This study consists of hierarchical Markov model along with quadtree structures in order to model the necessary information present in various special scales and a planar Markov model to tackle contextual spatial information at each resolution as well as the ensemble of causal decision trees for pixelwise modeling. Further reading on the methods for causal analysis used in machine learning can be found in the recent literature as given in [38]–[44].

Recursive partitioning models using gradient-based test statistics were considered in [45]. Several authors in [46]–[48] achieved the statistical stability by using a random forest resampling mechanism. Similarly, another study in [49] adopted a greedy and non-parametric regression technique, utilising gradient-based approximation.

Several studies [27], [46]–[48], [50]–[62] have considered the regression problems by using the random forest algorithms. Wager [21] proposed a method which is computationally efficient in generating generalised random forest (GRF). The estimates of this method are consistent and asymptotically normal, thereby providing a valid confidence interval. This method is designed to handle three main tasks, i.e., heterogeneous treatment effects, non-parametric quantile regression and conditional average treatment effects via instrumental variables.

The computational burden of trees ensemble could also be decreased without compromising the prediction accuracy by combining a small number of diverse and accurate trees. This can be achieved by using out-of-bag prediction errors from each training bootstrap sample in order to select the optimal trees on the basis of their individual performance [63]. The proposed method is a modified version of the generalised random forest [21], which involves generating a large number of causal trees and then selecting a proportion of those trees whose error rate is minimum among all the constructed causal trees.

III. METHODS

This section provides a detailed description of the method used in this paper. Before introducing the proposed method, it is deemed important to introduce the causal decision tree as a building block of the suggested optimal causal trees ensemble (*OCTE*).

A. CAUSAL DECISION TREE

Using the decision tree for estimating the heterogeneous treatment effect is totally different from the classical decision tree used for classification and regression problems. The classical decision tree uses a function mapping characteristics to the response variable about an individual. This can be illustrated from the tree given in Figure 1 as discussed in [64], where the matchmaking mobile app i.e. "Tinder" is used to cure a particular disease. The decision tree is unable to identify the true causal effect since majority of the young people use "Tinder" as compared to old people. Moreover, old people have little chance of recovery from the disease as compared with young people. Thus, the comparison between the two groups, which are not actually comparable, led to a misleading decision tree. A standard decision tree may thus be considered as an appropriate choice in terms of predicting the recovery from the disease but fails to identify the true cause of the disease. Furthermore, its nodes posses no causal interpretation.

On the other hand, the causal decision tree calculates the average of the treated and untreated observations in each node and then computes the difference between these averages, which represents the actual treatment effect in that node.

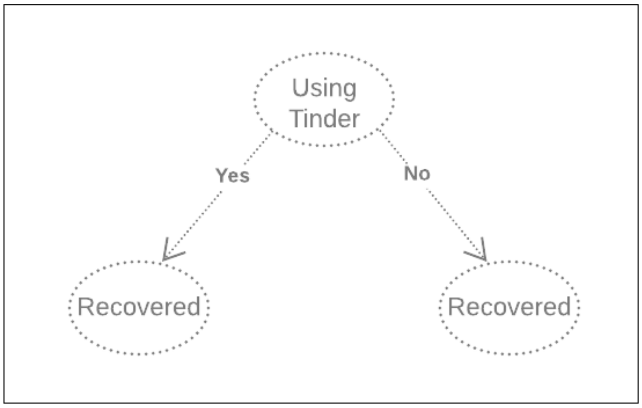

FIGURE 1. A standard classification tree.

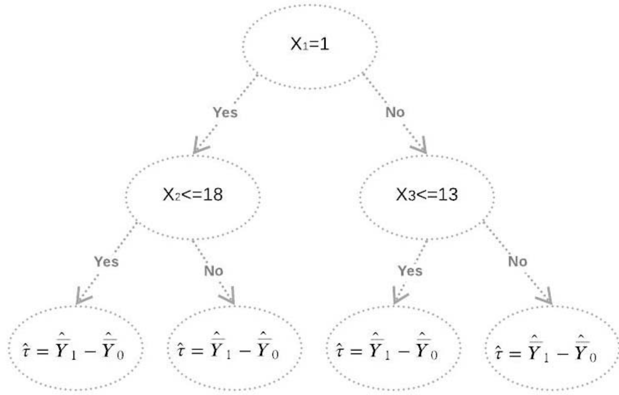

FIGURE 2. A standard causal decision tree.

Estimating the individual treatment effect, i.e. $\tau_i = Y_{1i} - Y_{0i}$, is not possible in real world problems, because the outcome of the ith individual is either Y_{1i} (the sample is treated) or Y_{0i} (the sample is untreated). One of the two outcomes, i.e. Y_{1i} or Y_{0i}, has to be predicted, using the counterfactual model (the potential outcome model). For example, suppose an individual has salary Y_{0i} and education below secondary level (untreated). We want to know the ith individual's salary Y_{1i} if the person had education above secondary level (treated). The average treatment effect of a group (population) is simply the average of the individual treatment effects included in the population, i.e.,

$$E[\tau_i] = E[Y_{1i}] - E[Y_{0i}].$$

A general work flow of a causal decision tree is given in Figure 2.

Decision trees for causal inference are generally used to separate data into buckets in order to estimate the average treatment effects within each node. The process of decision tree learning for causal inference can be separated into two steps for each of these tasks, commonly referred to as the splitting step and the estimation step, respectively. Therefore, a causal decision tree can effectively be used for estimating heterogeneous treatment effects in a computationally efficient manner. However, its prediction in the cases of regression and classification problems is less efficient.

Although a single causal decision tree model is interpretable and fast, the estimates it returns for heterogeneous treatment effects might not be generalisable. Therefore, the ensemble of causal decision trees can solve this problem by providing robust estimates for causal relationships at the cost of interpretability without significantly increasing computational cost. Combining a few accurate and diverse causal decision trees could provide improved estimates and might be taken forward in the direction of improving interpretability of the standard causal decision trees ensemble.

This work aims at improving the causal random forest with the help of best trees selection for size reduction and improved estimation. To achieve this, B sub-samples are taken from the given training data $L = (X, Y, W)$, where X is the feature space, Y is the response and W is the binary treatment. A causal decision tree is grown on each sub-sample. The performance of trees is evaluated on the basis of out-of-sample observations and ranked accordingly. Trees having the smallest error estimates on the out-of-sample observations are selected, while the rest of trees are discarded. Then, the selected trees are combined to form the optimal causal tree ensemble.

Partitioning of the given training data $L = (X, Y, W)$ is carried out randomly into two non-overlapping groups, i.e., $L_b = (X_b, Y_b, W_b)$ and $L_v = (X_v, Y_v, W_v)$. Then causal trees are grown, each on a sub-sample from $L_b = (X_b, Y_b, W_b)$. While doing so, a random subset of $p' < p$ features is selected from the entire set of features at each node of the causal tree; (where, p is total number of features and p' is a subset of features taken from total p features randomly). This will add additional randomness to the causal trees. As the observations in $L_v = (X_v, Y_v)$ take no part in the training phase of the causal trees, error estimates are calculated for these out-of-sample observations and the trees are ranked in descending order with respect to the error estimates. The error estimates taken in this paper are the standard errors for out-of-sample observations by each tree. The final ensemble is constructed by choosing the top ranked M causal trees.

B. STEPS OF THE PROPOSED (OCTE)

The proposed algorithm considers the following steps to assess treatment effect:

1) Select B number of sub-samples from the training part of the dataset i.e. $L = (X, Y, W)$.
2) Use generalized random forest for growing a causal decision tree on each sub-sample.
3) Arrange the causal trees according to their out-of-sample predicted standard errors.
4) The best M causal trees are chosen having the smallest individual prediction standard error on out-of-sample observations.
5) Combine the M selected trees to form an optimal causal decision trees forest and use it to predict the treatment effect of new/test data points.

Algorithm 1 Pseudocode of the Proposed *OCTE*

B: Number of sub samples taken from training data i.e. $L = (X, Y, W)$.
M: Number top ranked causal trees having minimum out of sample standard error.
for $t = 1 \to B$ **do**
    Grow a causal decision tree using generalized random forest;
    Compute error rate on out-of-sample observations;
    Save all the trees;
    Save the out-of-sample errors;
end for
Rank the causal trees based on the out-of-sample standard error;
Select top ranked M causal trees;
Combine the M selected causal trees to construct optimal causal tree ensemble (*OCTE*);
Use *OCTE* for estimating treatment effect and predicting unseen data.

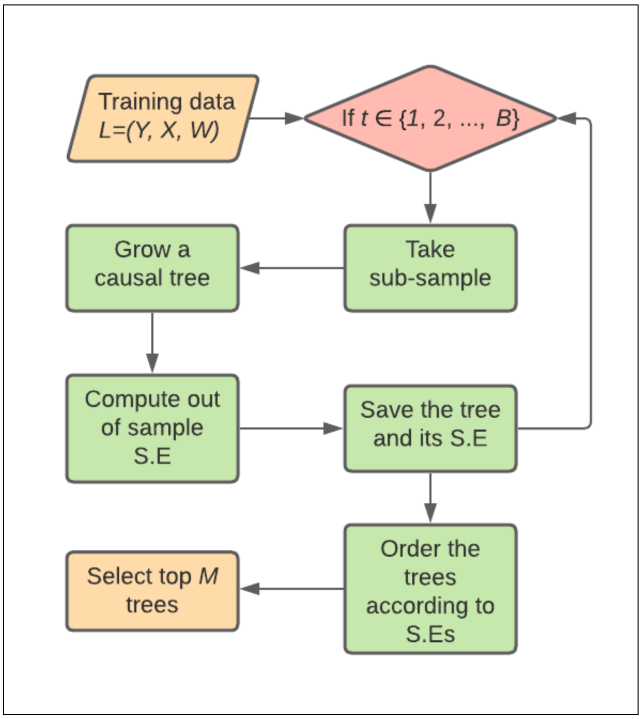

FIGURE 3. Flow chart of the proposed "OCTE" procedure.

Pseudocode of the proposed method *OCTE* is given in Algorithm 1 along with an illustrating flow chart in Figure 3.

IV. EXPERIMENTS AND RESULTS

In this paper, the proposed *OCTE* is assessed using five different simulation scenarios. It is then compared with five state-of-the-art methods, i.e., OLS, LASSO, Ridge, causal tree and causal random forest.

The *OCTE* is also applied on a real dataset, the nationally representative Labor Force Survey of Pakistan (LFSP). The LFSP data include records from 2017 to 2018 taken from

Pakistan's Bureau of Statistics. The Labor force survey is a nationwide survey containing micro-data from all over the country's demographic and employment information.

A. SIMULATED DATA

***Design 1:* Normally Distributed U_j and Linear Outcome Model**

The simulated data models are based on [65]–[68].

1) Each cluster consists of n_j observations, where $j = 1, 2, \ldots, J$, (where, J is the total number of clusters). Each cluster is generated by drawing a random number from a normal distribution with a small standard deviation and rounded mean I to the closest integer. All generated data contain about 4, 000 observations, and a conditional sample size of five instances is used for (J, I).
2) For all observations i.e., $i = 1, 2, \ldots, n_j$ for jth cluster, simulate confounders with individual level $X_{ij} = (X_{1ij}, X_{2ij}, X_{3ij})$, measured confounder for cluster level Z_j, and unmeasured confounder for cluster level U_j i.e.

$$\begin{aligned} X_{1ij} &\sim Unif(-1, 1), X_{2ij} \sim N(0, 1), \\ X_{3ij} &\sim Unif(0, 1), \\ Z_j &\sim Unif(-1, 1) \text{ and } U_j \sim N(0, 1). \end{aligned}$$

3) Status of individual treatment W_{ij} is generated from the logistic model propensity score as follow.

$$\begin{aligned} logit(e_{ij}) = &-0.6 + 0.3X_{1ij} + 0.3X_{2ij} + 0.3X_{3ij} \\ &+0.3Z_j + 0.4X_{1ij}^2 + 0.4X_{2ij}Z_j \\ &+0.4X_{1ij}I(X_{3ij} \leq 0.3) + 0.3Z_j, \end{aligned}$$

and $W_{ij} \sim Bernoulli(e_{ij})$.
4) The potential outcomes Y_{ij1}, Y_{ij0} and observed response Y_{ij} are generated from the regression model as follow.

$$\begin{aligned} Y_{ij}(w) = &-70 + 2X_{1ij} + 2X_{2ij} + 2X_{3ij} \\ &+2Z_j + 2X_{1ij}^2 + 2X_{2ij}Z_j \\ &+2X_{1ij}I(X_{3ij \leq 0.3}) + 2U_j \\ &+W(2 + 2X_{3ij} + 2Z_j) + r_{ij}, \\ Y_{ij} = &W_{ij}Y_{ij1} + (1 - W_{ij})Y_{ij0}, \end{aligned}$$

and $r_{ij} \sim N(0, 1)$, where r_{ij} is random error for ith sample in jth cluster.

***Design 2:* Uniformly Distributed U_j and Linear Outcome Model**

This design utilizes a similar model for data generating as Design 1, but the only difference is that U_j (unmeasured cluster level confounder) has uniform distribution i.e. $U_j \sim U(-2, 2)$.

***Design 3:* Uniformly Distributed U_j, Linear Outcome Model, and Misspecified Working Models**

This design is also similar to Design 2, but higher order terms in the outcome model are ignored.

***Design 4:* Uniformly Distributed U_j and Nonlinear Binary Outcome Model**

TABLE 1. Summary of Designs 1-5.

	Specification		
Design	Distribution	Outcome model	Estimated models with higher order terms
1	N(0, 1)	Linear	Yes
2	Unif(-2, 2)	Linear	Yes
3	Unif(-2, 2)	Linear	No
4	Unif(-2, 2)	Non-Linear	Yes
5	Unif(-2, 2)	Linear with exponential error term	Yes

This design is also similar to Design 2, but the difference is that the outcome model nonlinear and binary. The outcome model is given by

$$\begin{aligned} logit(P(Y_{ij}(w) = 1|X_{ij}, Z_j, U_j)) = &-0.6 + 0.3X_{1ij} \\ &+0.3X_{2ij} + 0.3X_{3ij} + 0.3Z_j + 0.3X_{1ij}^2 \\ &+0.3X_{2ij}^2Z_j + 0.3X_{1ij}I(X_{3ij} \leq 0.3) \\ &+0.3Z_j + \beta_1 U_j + \beta_2 w(0.5 + 0.3X_{3ij} \\ &+0.3Z_j + \beta_2 U_j), \end{aligned}$$

and $Y_{ij}(w) \sim Bernoulli(P(Y_{ij}(w) = 1|X_{ij}, Z_j, U_j)), Y_{ij} = W_{ij}Y_{ij1}(1 - W_{ij})Y_{ij0}$.

***Design 5:* Uniformly Distributed U_j and Linear Outcome with Exponential Error Model**

The construction of Design 5 is almost similar to Design 2 with U_j as uniformly distributed and linear outcome model.

All five designs consist of about 4000 samples and data in each design is divided into 70% training and 30% testing parts. *MSE*, *RMSE*, *MAD* and correlation coefficient (r) are used as performance measures for both conditional average treatment effect (*CATE*) i.e. τ_i as will as average treatment effect (*ATE*) i.e. τ.

For *CATE*, 500 realization are made under each design and *MSE*, *RMSE*, *MAD* and correlation coefficient (r) are calculated reporting their average values. Expressions of the metrics used are given bellow.

$$\begin{aligned} MSE &= \frac{1}{n}\sum\nolimits_{i=1}^{n}(\tau_i - \hat{\tau}_i)^2, \\ RMSE &= \sqrt{\frac{1}{n}\sum\nolimits_{i=1}^{n}(\tau_i - \hat{\tau}_i)^2}, \\ MAD &= \frac{1}{n}\sum\nolimits_{i=1}^{n}|\tau_i - \hat{\tau}_i|, \\ r &= \frac{\sum_{i=1}^{n}(\hat{\tau}_i - \bar{\hat{\tau}}_i)(\tau_i - \bar{\tau}_i)}{\sqrt{\sum_{i=1}^{n}(\hat{\tau}_i - \bar{\hat{\tau}}_i)^2}\sqrt{\sum_{i=1}^{n}(\tau_i - \bar{\tau}_i)^2}}. \end{aligned}$$

B. SIMULATED DATA RESULTS

Table 4 shows the results of the proposed method (*OCTE*) and all the other methods considered in this study, in terms of conditional average treatment effect (*CATE*). The results suggest that the proposed *OCTE* outperformed all the other stat-of-the-art procedure on almost all the five designs. The results are also shown in the form of bar plots in Figures 4-7.

TABLE 2. Execution time (in seconds) of *OCTE* and CF for different number of trees.

Number of trees	OCTE	CF
50	7.306	0.296
100	7.65	0.347
200	8.412	0.474
300	8.497	0.985
400	10.412	0.986
500	11.685	1.198
1000	17.622	2.309
2000	27.281	4.688

TABLE 3. Labour Force Survey of Pakistan (LFSP) data description.

Variable	Description
Y	Continuous outcome: Logged monthly income of workers.
W	Education: Used as binary treatment; 1 = Education above matric, 0 = Education bellow matric.
X_1	Experience: Experience is calculated using (Age-year of schooling-6).
X_2	Kind of Enterprise: A categorical variable with four groups, Government, Private, Public enterprise, other.
X_3	Industry: Industry has 22 categories of different industries, such as Manufacturing, Electricity, etc.
X_4	Hours worked: Hours worked is the sum of total hours an individual work.
X_5	Gender: Tow classes, male and female.
X_6	Marital status: Marital status is classified as widow, single, married, divorced.
X_7	Employment status: Categorical with four classes, paid employees, Employer, Self Employed, Contributing Family Helpers
X_8	Occupation: It is a categorical variable with nine classes and shows the occupation of employees, such as manager, professional, etc.
X_9	Training: Binary variable and shows whether training has been given to the employee for a job or not.
X_{10}	Region: Residence of employee, which has two categories, rural and urban.
X_{11}	Province: Province of residence, which has four categories i.e. Khyber Pakhtunkhwa, Punjab, Sindh and Balochistan.

The proposed *OCTE* provides minimum mean square error (*MSE*) on first four designs, while causal forest (CF) gives optimal value for Design 5. The OLS, LASSO, Ridge and CT did not perform well on any design. In terms of root mean square error (*RMSE*), the proposed *OCTE* performed better than the other methods on four designs and CF outperformed the others on Design 5. Apart from *OCTE* and CF, the remaining methods did not outperformed the rest of the methods on any of the designs. Similarly, *OCTE* is giving optimal results in terms of mean absolute deviation (*MAD*) as compared to the other methods. In terms of Pearson's product moment correlation coefficients (r), the *OCTE* also outperformed the other procedures on Designs 3 and has similar performance on Designs 1 and 4, while OLS outperformed the rest of the methods on Designs 4 and 5.

Table 5 shows the results for average treatment effect (ATE) for the 5 scenarios. It is evident from the results that the proposed OCTE had better achievements than the other methods in four scenarios in terms of *MSE*, *RMSE* and *MAD*, and it outperformed them in three scenarios in terms of correlation. The CF method, on the other hand, gave the same correlation value in Design 3 as that of OCTE and outperformed the other methods in Design 4. However,

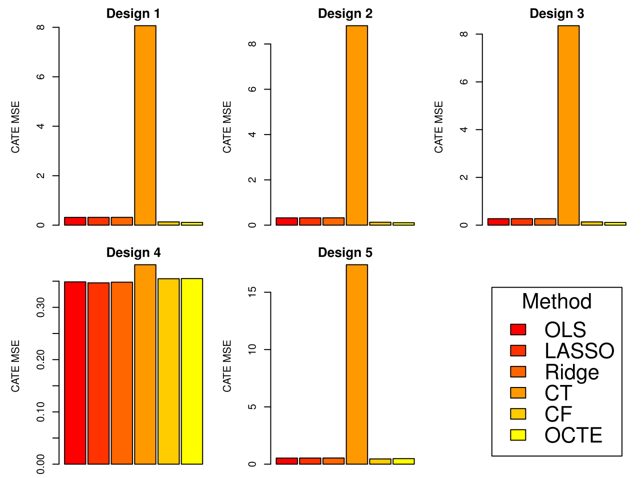

FIGURE 4. Bar plots of *MSE* computed for CATE.

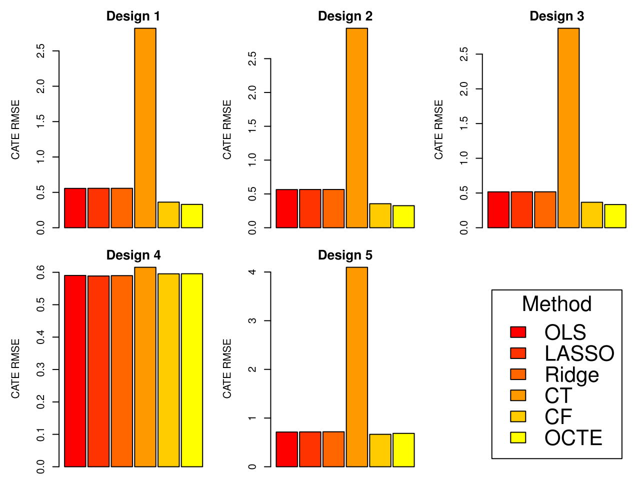

FIGURE 5. Bar plots of *RMSE* computed for CATE.

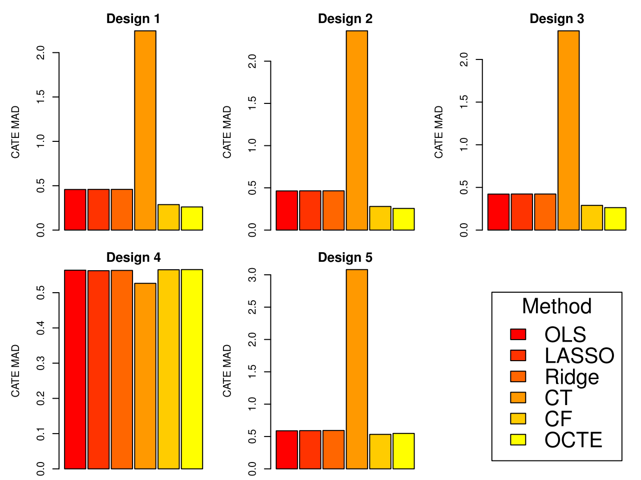

FIGURE 6. Bar plots of *MAD* computed for CATE.

LASSO had the best achievement in Design 5 in terms of correlation. For a visual illustration, the results are also shown in the form of bar plots in Figures 12-15. The results obtained in Table 4 and Table 5 can also be seen in the form of bar-plots in Figures 4-7 and Figures 12-15, respectively.

TABLE 4. Conditional average treatment effect results of all the methods.

Design	Methods	*MSE*	*RMSE*	*MAD*	*r*
	OLS Regression	0.31432	0.55620	0.45744	**0.97429**
	LASSO Regression	0.31560	0.55735	0.45846	0.97426
1	Ridge Regression	0.31556	0.55732	0.4585	0.97421
	Causal Tree	8.06559	2.81935	2.2457	0.51432
	Causal Forest	0.13283	0.36201	0.28706	0.97146
	OCTE	**0.11038**	**0.32978**	**0.26069**	**0.97429**
	OLS Regression	0.32334	0.56417	0.46353	0.97274
	LASSO Regression	0.32448	0.56518	0.46445	0.97271
2	Ridge Regression	0.32475	0.56542	0.46473	0.97267
	Causal Tree	8.80697	2.94933	2.35961	0.45639
	Causal Forest	0.12809	0.35515	0.27959	0.97215
	OCTE	**0.10797**	**0.32611**	**0.25628**	**0.97405**
	OLS Regression	0.27129	0.51680	0.42195	**0.97447**
	LASSO Regression	0.27234	0.51780	0.42278	0.97442
3	Ridge Regression	0.27203	0.51750	0.42259	0.97441
	Causal Tree	8.35285	2.87226	2.33565	0.47505
	Causal Forest	0.13646	0.36671	0.28950	0.97115
	OCTE	**0.11317**	**0.33371**	**0.26329**	**0.97447**
	OLS Regression	0.31432	0.55620	0.45744	**0.97429**
	LASSO Regression	0.31560	0.55735	0.45846	0.97426
4	Ridge Regression	0.31556	0.55732	0.45850	0.97421
	Causal Tree	8.06559	2.81935	2.24570	0.51432
	Causal Forest	0.13283	0.36201	0.28706	0.97146
	OCTE	**0.11038**	**0.32978**	**0.26069**	0.97394
	OLS Regression	0.53181	0.71310	0.58835	**0.95295**
	LASSO Regression	0.53532	0.71556	0.59051	0.95268
5	Ridge Regression	0.53923	0.71823	0.59327	0.95260
	Causal Tree	17.4036	4.09703	3.08054	0.32028
	Causal Forest	**0.45783**	**0.66611**	**0.53278**	0.88576
	OCTE	0.48421	0.68542	0.54742	0.87311

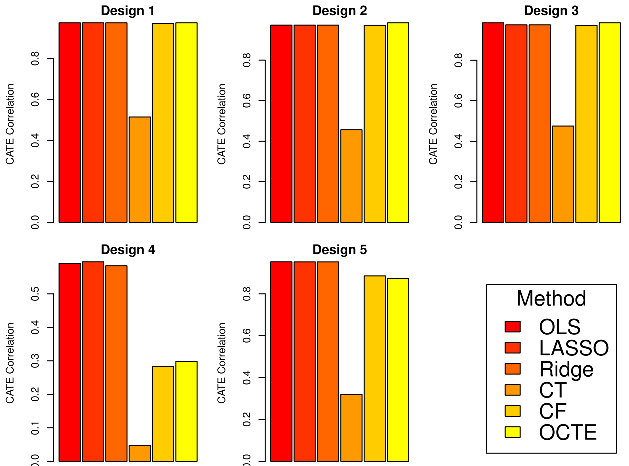

FIGURE 7. Bar plots of *r* computed for CATE.

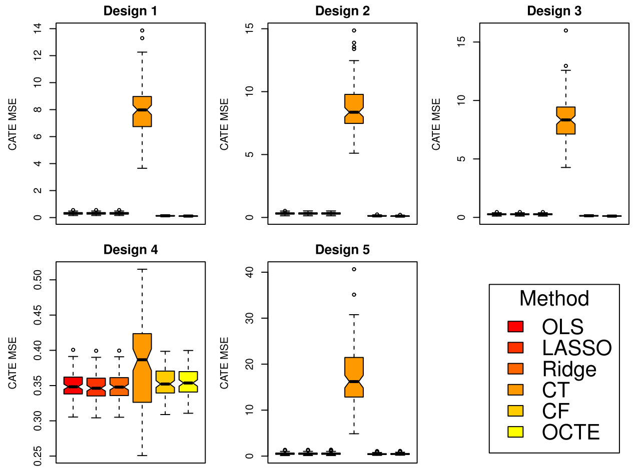

FIGURE 8. Boxplots of *MSE* computed for CATE.

For further assessment, boxplots of *MSE*, *RMSE*, *MAD* and correlation values of *CATE* obtained from all the 500 runs of the simulated data for each design are displayed in Figures 8, 9, 10 and 11. The boxplots reveal that the *OCTE* is more consistent in comparison with the other well-known methods.

Moreover, the execution or running times (in seconds) of the new method and the causal forest method (CF) are also given in Table 2, where it can be noticed that, as the number of trees grow, the execution increases linearly, i.e. $f(B) \leq C \times O(B)$, where B is the number of trees. The execution time of the *OCTE* is greater than that of the CF

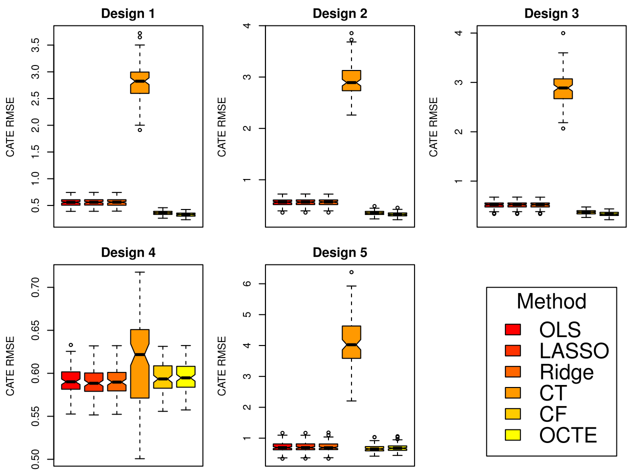

FIGURE 9. Boxplots of *RMSE* computed for CATE.

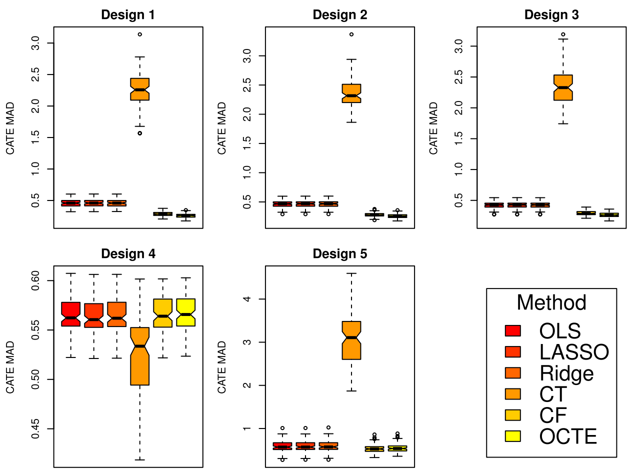

FIGURE 10. Boxplots of *MAD* computed for CATE.

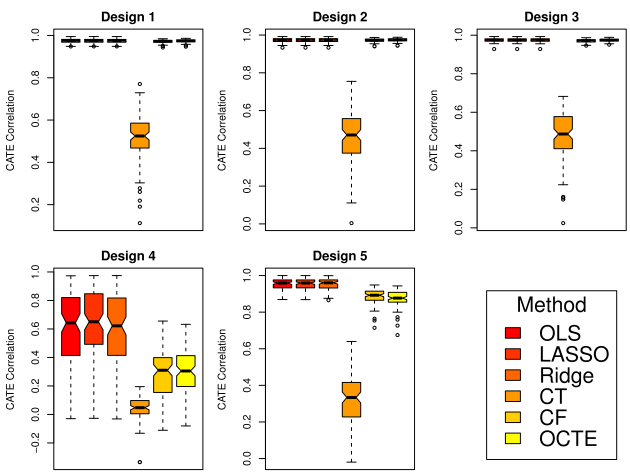

FIGURE 11. Boxplots of *r* computed for CATE.

method, due to the additional tree selection step. To reduce the execution time, Step 3 of the proposed algorithm can be parallelised using existing tools, such as the "parallel" R package [69].

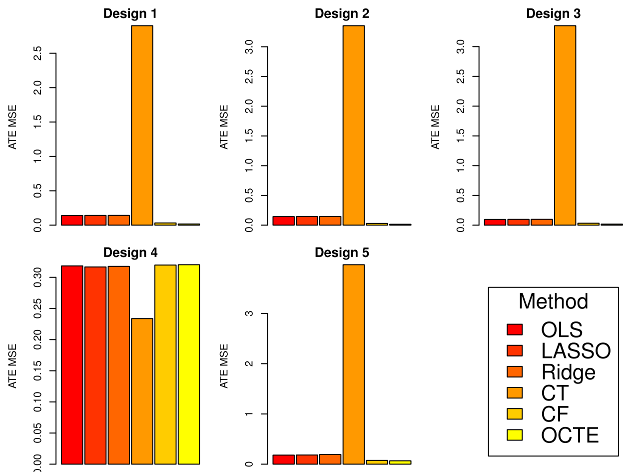

FIGURE 12. Bar plots of *MSE* computed for ATE.

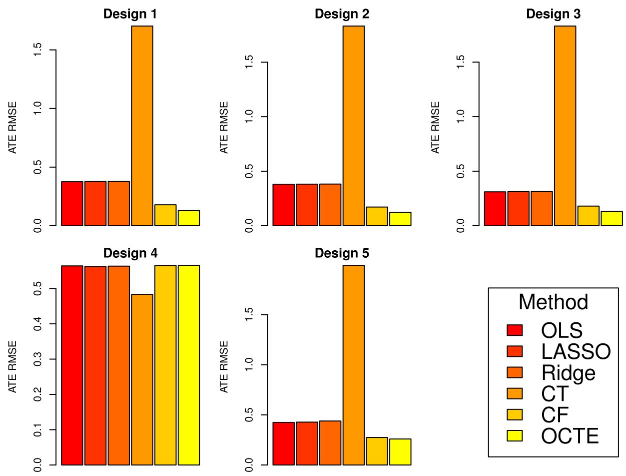

FIGURE 13. Bar plots of *RMSE* computed for ATE.

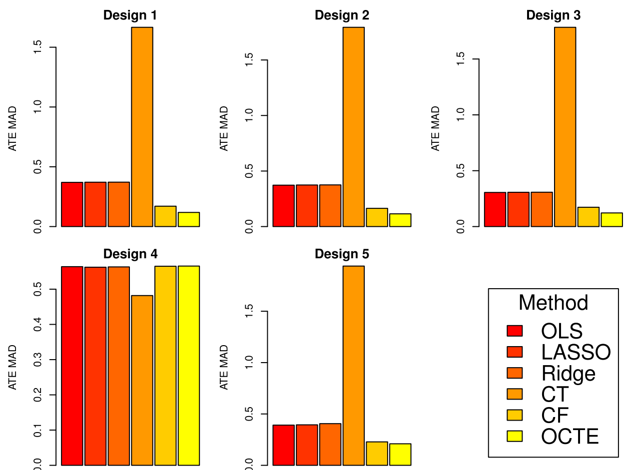

FIGURE 14. Bar plots of *MAD* computed for ATE.

C. LABOUR FORCE SURVEY OF PAKISTAN (LFSP) AS AN EXAMPLE

In this research, the nationally representative labor force survey of Pakistan (LFSP) data from 2017 to 2018 is taken from

TABLE 5. Average treatment effect results of all the methods.

Design	Methods	*MSE*	*RMSE*	*MAD*	*r*
1	OLS Regression	0.14080	0.37524	0.36957	0.77690
	LASSO Regression	0.14187	0.37666	0.37101	0.77718
	Ridge Regression	0.14224	0.37715	0.37150	0.77667
	Causal Tree	2.90104	1.70325	1.66662	0.45533
	Causal Forest	0.03190	0.17860	0.17043	0.85308
	OCTE	**0.01665**	**0.12902**	**0.11817**	**0.85886**
2	OLS Regression	0.14431	0.37989	0.37279	0.83500
	LASSO Regression	0.14545	0.38135	0.37425	0.83498
	Ridge Regression	0.14608	0.38221	0.37513	0.83415
	Causal Tree	3.35401	1.83139	1.79410	0.48703
	Causal Forest	0.02936	0.17133	0.16403	0.90414
	OCTE	**0.01512**	**0.12297**	**0.11456**	**0.90952**
3	OLS Regression	0.09665	0.31088	0.30538	0.85601
	LASSO Regression	0.09756	0.31234	0.30685	0.85570
	Ridge Regression	0.09794	0.31295	0.30747	0.85629
	Causal Tree	3.35188	1.83081	1.78372	0.23637
	Causal Forest	0.03239	0.17997	0.17289	**0.89044**
	OCTE	**0.01725**	**0.13135**	**0.12203**	**0.89044**
4	OLS Regression	0.31838	0.56425	0.56392	-0.01819
	LASSO Regression	0.31669	0.56275	0.56240	-0.05131
	Ridge Regression	0.31762	0.56358	0.56323	-0.05942
	Causal Tree	**0.23373**	**0.48346**	**0.48189**	-0.16226
	Causal Forest	0.31973	0.56544	0.56512	**-0.01687**
	OCTE	0.32029	0.56594	0.56562	-0.03722
5	OLS Regression	0.18050	0.42486	0.39127	0.45418
	LASSO Regression	0.18270	0.42744	0.39414	**0.45434**
	Ridge Regression	0.19254	0.43880	0.40576	0.44677
	Causal Tree	3.96988	1.99246	1.93983	0.24032
	Causal Forest	0.07522	0.27426	0.22826	0.40453
	OCTE	**0.06731**	**0.25944**	**0.20970**	0.37665

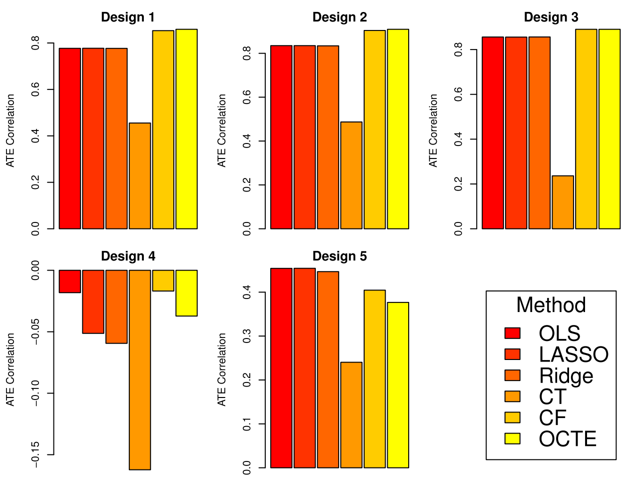

FIGURE 15. Bar plots of *r* computed for ATE.

Pakistan Bureau of Statistics. LFSP is a nationwide survey consisting of labours employment information from all over the country at micro level.

The working sample used is based on those in wage employment and comprises a total of 272610 workers. The analysis is restricted to those older than 10. Missing values and unusable observations are discarded, leaving a total of 27964 observations. The variables used to analyse each

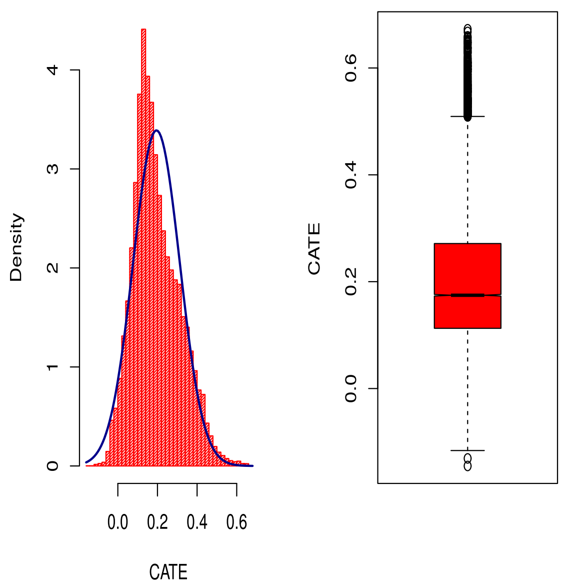

FIGURE 16. CATE computed by OCTE for LFSP Data.

worker's wage include hours worked, education, occupation, residence (in urban or rural area and in one of the four provinces), schooling attainment, gender, employment status,

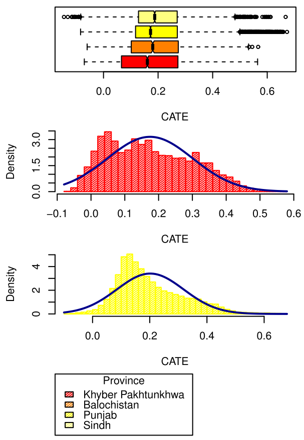

FIGURE 17. Heterogeneity based on Provinces for LFSP Data.

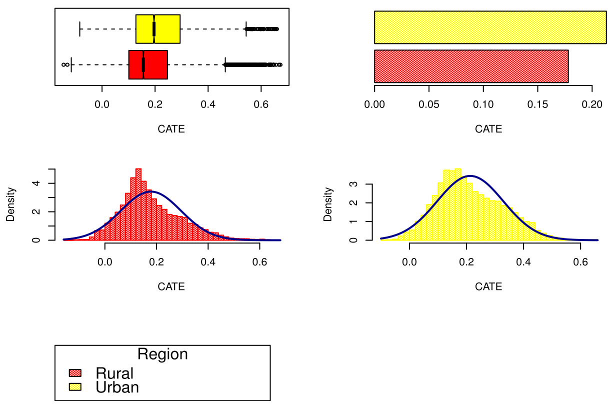

FIGURE 18. Heterogeneity based on Region for LFSP Data.

marital status, experience, industry, kind of enterprise and training. A brief description of the variables is given in Table 3.

D. LFSP DATA RESULTS

Figure 16, shows the conditional average treatment effect (*CATE*) of the proposed *OCTE* computed for *LFSP* data. It can be observed from the figure that the treatment (education) has a positive effect on the average income of the individuals. This implies that higher education leads to higher income of the individuals. Similar conclusion could be drawn from the box-plot constructed in the same figure.

Figures 17-19 discuss the heterogeneity of education in the variables "Province", "Region" and "Gender", respectively. From Figure 17, it is clear that in province Khyber Pakhtunkhwa, the average effect of education on the income of the individuals is less than the rest of the provinces. Punjab and Sindh posses almost equal average effects of education on the income of the individuals. Figure 18 indicates the heterogeneity of education between the rural and urban areas where it is clear that in rural areas, the education has less

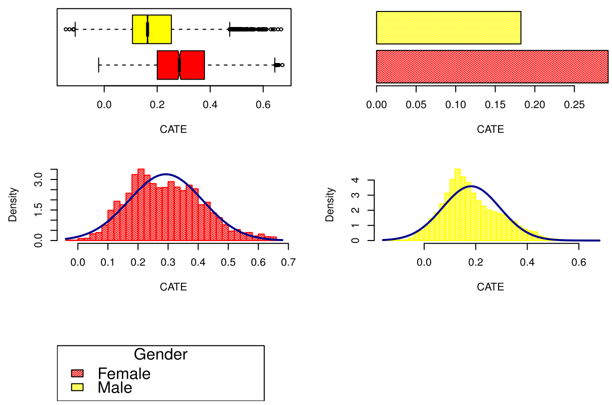

FIGURE 19. Heterogeneity based on Gender for LFSP Data.

effect on the income of the individuals. In case of variable "Gender", education has minimum effect on the income of males as compared to females (Figure 19). In a nut shell, education is observed to have heterogeneous effect among the variables i.e. gender and region, whereas in variable province it has an approximately homogeneous effect.

V. CONCLUSION

This research proposed a causal tree selection method based on out-of-sample standard error. The procedure grows a large number of causal trees, each on a random sub-sample taken from the training data. The proposed method estimates standard errors of all the trees based on out-of-sample observations. The causal trees grown are ranked with respect to the standard errors, and the top M trees are selected. The top ranked trees are combined for the final ensemble. The novel *OCTE* is assessed based on simulated data, generated under five different designs and compared with common procedures, OLS, LASSO, ridge, causal tree and causal decision trees forest. For assessing the proposed *OCTE*, performance metrics *MSE*, *RMSE*, *MAD* and Pearson's correlation coefficient (r) are used. In general, the proposed algorithm outperformed the rest of the methods in almost all the cases.

The *OCTE* method demonstrated improved prediction performance compared to the rest of the methods considered in the paper. Moreover, the method is effective in estimating heterogeneous causal treatment effects. Since the method uses only few accurate causal trees, the idea could be further be extended to mitigate the interpretability issue of the standard causal trees forest.

REFERENCES

[1] J. E. Brand, J. Xu, B. Koch, and P. Geraldo, "Uncovering sociological effect heterogeneity using tree-based machine learning," *Sociol. Methodol.*, vol. 51, no. 2, pp. 189–223, Aug. 2021, Art. no. 0081175021993503.

[2] R. E. Schapire, "The strength of weak learnability," *Mach. Learn.*, vol. 5, no. 2, pp. 197–227, Jun. 1990.

[3] P. Domingos, "Using partitioning to speed up specific-to-general rule induction," in *Proc. AAAI Workshop Integrating Multiple Learned Models*. Princeton, NJ, USA: Citeseer, 1996, pp. 29–34.

[4] T. Mitchell, *Machine Learning*. Burr Ridge, IL, USA: McGraw-Hill, 1997.
[5] K. Turner and J. Ghosh, "Error correlation and error reduction in ensemble classifiers," *Connection Sci.*, vol. 8, no. 3, pp. 385–404, 1996.
[6] K. M. Ali and M. J. Pazzani, "Error reduction through learning multiple descriptions," *Mach. Learn.*, vol. 24, no. 3, pp. 173–202, Sep. 1996.
[7] S. Bernard, L. Heutte, and S. Adam, "On the selection of decision trees in random forests," in *Proc. Int. Joint Conf. Neural Netw.*, Jun. 2009, pp. 302–307.
[8] N. Meinshausen, "Node harvest," *Ann. Appl. Statist.*, vol. 4, no. 4, pp. 2049–2072, Dec. 2010.
[9] T. M. Oshiro, P. S. Perez, and J. A. Baranauskas, "How many trees in a random forest? in *Proc. Int. Workshop Mach. Learn. Data Mining Pattern Recognit.* Berlin, Germany: Springer, 2012, pp. 154–168.
[10] P. Latinne, O. Debeir, and C. Decaestecker, "Limiting the number of trees in random forests," in *Proc. Int. Workshop Multiple Classifier Syst.* Berlin, Germany: Springer, 2001, pp. 178–187.
[11] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[12] J. Fan, M. Farmen, and I. Gijbels, "Local maximum likelihood estimation and inference," *J. Roy. Stat. Soc. B, Stat. Methodol.*, vol. 60, no. 3, pp. 591–608, 1998.
[13] W. K. Newey, "Kernel estimation of partial means and a general variance estimator," *Econ. Theory*, vol. 10, no. 2, pp. 1–21, Jun. 1994.
[14] J. G. Staniswalis, "The kernel estimate of a regression function in likelihood-based models," *J. Amer. Stat. Assoc.*, vol. 84, no. 405, pp. 276–283, Mar. 1989.
[15] C. J. Stone, "Consistent nonparametric regression," *Ann. Statist.*, vol. 5, no. 4, pp. 595–620, Jul. 1977.
[16] R. Tibshirani and T. Hastie, "Local likelihood estimation," *J. Amer. Stat. Assoc.*, vol. 82, no. 398, pp. 559–567, 1987.
[17] A. Lewbel, "A local generalized method of moments estimator," *Econ. Lett.*, vol. 94, no. 1, pp. 124–128, Jan. 2007.
[18] B. E. Honoré and E. Kyriazidou, "Panel data discrete choice models with lagged dependent variables," *Econometrica*, vol. 68, no. 4, pp. 839–874, Jul. 2000.
[19] L. Su, I. Murtazashvili, and A. Ullah, "Local linear GMM estimation of functional coefficient IV models with an application to estimating the rate of return to schooling," *J. Bus. Econ. Statist.*, vol. 31, no. 2, pp. 184–207, Apr. 2013.
[20] J. M. Robins and Y. Ritov, "Toward a curse of dimensionality appropriate (CODA) asymptotic theory for semi-parametric models," *Statist. Med.*, vol. 16, no. 3, pp. 285–319, Feb. 1997.
[21] S. Athey, J. Tibshirani, and S. Wager, "Generalized random forests," *Ann. Statist.*, vol. 47, no. 2, pp. 1148–1178, Apr. 2019.
[22] Y. Amit and D. Geman, "Shape quantization and recognition with randomized trees," *Neural Comput.*, vol. 9, no. 7, pp. 1545–1588, 1997.
[23] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, 1996.
[24] T. G. Dietterich, "An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization," *Mach. Learn.*, vol. 40, no. 2, pp. 139–157, 2000.
[25] T. K. Ho, "The random subspace method for constructing decision forests," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 8, pp. 832–844, Aug. 1998.
[26] T. Hothorn, B. Lausen, A. Benner, and M. Radespiel-Tröger, "Bagging survival trees," *Statist. Med.*, vol. 23, no. 1, pp. 77–91, Jan. 2004.
[27] N. Meinshausen and G. Ridgeway, "Quantile regression forests," *J. Mach. Learn. Res.*, vol. 7, no. 6, pp. 983–999, 2006.
[28] D. W. K. Andrews, "Tests for parameter instability and structural change with unknown change point," *Econometrica, J. Econ. Soc.*, vol. 61, no. 4, pp. 821–856, 1993.
[29] B. E. Hansen, "Testing for parameter instability in linear models," *J. Policy Model.*, vol. 14, no. 4, pp. 517–533, Aug. 1992.
[30] N. L. Hjort and A. Koning, "Tests for constancy of model parameters over time," *J. Nonparametric Statist.*, vol. 14, nos. 1–2, pp. 113–132, Jan. 2002.
[31] J. Nyblom, "Testing for the constancy of parameters over time," *J. Amer. Stat. Assoc.*, vol. 84, no. 405, pp. 223–230, Mar. 1989.
[32] A. Zeileis, "A unified approach to structural change tests based on ML scores, *F* statistics, and OLS residuals," *Econ. Rev.*, vol. 24, no. 4, pp. 445–466, Oct. 2005.
[33] A. Zeileis and K. Hornik, "Generalized M-fluctuation tests for parameter instability," *Statistica Neerlandica*, vol. 61, no. 4, pp. 488–508, Nov. 2007.
[34] W. Ploberger and W. Krämer, "The CUSUM test with OLS residuals," *Econometrica, J. Econ. Soc.*, vol. 60, no. 2, pp. 271–285, 1992.
[35] G. Bakirli and D. Birant, "DTreeSim: A new approach to compute decision tree similarity using re-mining," *TURKISH J. Electr. Eng. Comput. Sci.*, vol. 25, pp. 108–125, 2017.
[36] Z. Hu, F. Li, X. Wang, and Q. Lin, "Description length guided unified Granger causality analysis," *IEEE Access*, vol. 9, pp. 13704–13716, 2021.
[37] M. Pastorino, A. Montaldo, L. Fronda, I. Hedhli, G. Moser, S. B. Serpico, and J. Zerubia, "Multisensor and multiresolution remote sensing image classification through a causal hierarchical Markov framework and decision tree ensembles," *Remote Sens.*, vol. 13, no. 5, p. 849, Feb. 2021.
[38] F. Yan, J. Ma, M. Li, R. Niu, and T. Tang, "An automated accident causal scenario identification method for fully automatic operation system based on STPA," *IEEE Access*, vol. 9, pp. 11051–11064, 2021.
[39] A. J. P. N. Palacios, J. E. N. Ricardo, I. A. C. Piza, and M. E. E. N. Herrería, "Phenomenological hermeneutical method and neutrosophic cognitive maps in the causal analysis of transgressions against the homeless," *Neutrosophic sets Syst.*, vol. 44, no. 1, p. 18, 2021.
[40] S. Singla, S. Wallace, S. Triantafillou, and K. Batmanghelich, "Using causal analysis for conceptual deep learning explanation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* USA: Inst. of Mathematical Statistics, 2021, pp. 519–528.
[41] J. Yoon, "Forecasting of real GDP growth using machine learning models: Gradient boosting and random forest approach," *Comput. Econ.*, vol. 57, no. 1, pp. 247–265, Jan. 2021.
[42] M. Pellagatti, C. Masci, F. Ieva, and A. M. Paganoni, "Generalized mixed-effects random forest: A flexible approach to predict university student dropout," *Stat. Anal. Data Mining, ASA Data Sci. J.*, vol. 14, no. 3, pp. 241–257, Jun. 2021.
[43] S. Zeng, M. A. Bayir, J. J. Pfeiffer, III, D. Charles, and E. Kiciman, "Causal transfer random forest: Combining logged data and randomized experiments for robust prediction," in *Proc. 14th ACM Int. Conf. Web Search Data Mining*, Mar. 2021, pp. 211–219.
[44] Y. Küçük, T. A. D. Henderson, and A. Podgurski, "Improving fault localization by integrating value and predicate based causal inference techniques," in *Proc. IEEE/ACM 43rd Int. Conf. Softw. Eng. (ICSE)*, May 2021, pp. 649–660.
[45] A. Zeileis, T. Hothorn, and K. Hornik, "Model-based recursive partitioning," *J. Comput. Graph. Statist.*, vol. 17, no. 2, pp. 492–514, 2008.
[46] L. Mentch and G. Hooker, "Quantifying uncertainty in random forests via confidence intervals and hypothesis tests," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 841–881, 2016.
[47] E. Scornet, G. Biau, and J.-P. Vert, "Consistency of random forests," *Ann. Statist.*, vol. 43, no. 4, pp. 1716–1741, Aug. 2015.
[48] S. Wager and S. Athey, "Estimation and inference of heterogeneous treatment effects using random forests," *J. Amer. Stat. Assoc.*, vol. 113, no. 523, pp. 1228–1242, Jul. 2018.
[49] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001.
[50] S. Arlot and R. Genuer, "Analysis of purely random forests bias," 2014, *arXiv:1407.3939*.
[51] Y. Lin and Y. Jeon, "Random forests and adaptive nearest neighbors," *J. Amer. Statist. Assoc.*, vol. 101, no. 474, pp. 578–590, 2006.
[52] G. Biau, "Analysis of a random forests model," *J. Mach. Learn. Res.*, vol. 13, pp. 1063–1095, Apr. 2012.
[53] G. Biau, L. Devroye, and G. Lugosi, "Consistency of random forests and other averaging classifiers," *J. Mach. Learn. Res.*, vol. 9, no. 9, pp. 2015–2033, 2008.
[54] G. Biau and E. Scornet, "A random forest guided tour," *Test*, vol. 25, no. 2, pp. 197–227, Jun. 2016.
[55] P. Bühlmann and B. Yu, "Analyzing bagging," *Ann. Statist.*, vol. 30, no. 4, pp. 927–961, 2002.
[56] M. Denil, D. Matheson, and N. D. Freitas, "Narrowing the gap: Random forests in theory and in practice," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 665–673.
[57] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Mach. Learn.*, vol. 63, no. 1, pp. 3–42, 2006.
[58] H. Ishwaran and U. B. Kogalur, "Consistency of random survival forests," *Statist. Probab. Lett.*, vol. 80, nos. 13–14, pp. 1056–1064, 2010.
[59] J. Sexton and P. Laake, "Standard errors for bagged and random forest estimators," *Comput. Statist. Data Anal.*, vol. 53, no. 3, pp. 801–811, Jan. 2009.
[60] S. Wager and G. Walther, "Adaptive concentration of regression trees, with application to random forests," 2015, *arXiv:1503.06388*.
[61] R. Zhu, D. Zeng, and M. R. Kosorok, "Reinforcement learning trees," *J. Amer. Stat. Assoc.*, vol. 110, no. 512, pp. 1770–1784, 2015.

[62] G. Fang, P. Xu, and W. Liu, "Automated ischemic stroke subtyping based on machine learning approach," *IEEE Access*, vol. 8, pp. 118426–118432, 2020.
[63] Z. Khan, A. Gul, A. Perperoglou, M. Miftahuddin, O. Mahmoud, W. Adler, and B. Lausen, "Ensemble of optimal trees, random forest and random projection ensemble classification," *Adv. Data Anal. Classification*, vol. 14, no. 1, pp. 97–116, Mar. 2020.
[64] J. Li, S. Ma, T. Le, L. Liu, and J. Liu, "Causal decision trees," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 2, pp. 257–271, Feb. 2017.
[65] M. Pastorino, A. Montaldo, L. Fronda, I. Hedhli, G. Moser, S. B. Serpico, and J. Zerubia, "Multisensor and multiresolution remote sensing image classification through a causal hierarchical Markov framework and decision tree ensembles," *Remote Sens.*, vol. 13, no. 5, pp. 849–873, 2021.
[66] F. Li, A. M. Zaslavsky, and M. B. Landrum, "Propensity score weighting with multilevel data," *Statist. Med.*, vol. 32, no. 19, pp. 3373–3387, Aug. 2013.
[67] M. S. Schuler, W. Chu, and D. Coffman, "Propensity score weighting for a continuous exposure with multilevel data," *Health Services Outcomes Res. Methodol.*, vol. 16, no. 4, pp. 271–292, Dec. 2016.
[68] J. Zetterqvist, S. Vansteelandt, Y. Pawitan, and A. Sjölander, "Doubly robust methods for handling confounding by cluster," *Biostatistics*, vol. 17, no. 2, pp. 264–276, Apr. 2016.
[69] *R: A Language and Environment for Statistical Computing*, R Core Team, R Found. Stat. Comput., Vienna, Austria, 2021.

NEELAM YOUNAS received the master's degree in statistics from Quaid-i-Azam University, Isalamabad, Pakistan in 2007, the M.Phil. degree in econometrics from the Pakistan Institute of Development of Economics (PIDE), Pakistan, in 2013, where she is currently pursuing the Ph.D. degree in econometrics. She has also worked as a Lecturer at King Saud University, Saudi Arabia, from 2009 to 2012, and for one year at Al Yammamah University, Saudi Arabia. She has been working with NUML and Bahria University Islamabad, as a Visiting Faulty, since January 2021. Her research interests include machine learning, applied econometrics, and micro econometrics.

AMJAD ALI received the bachelor's and master's degrees in statistics from the University of Peshawar, Pakistan, in 2014 and 2017, respectively, and the M.Phil. degree in data science from the Department of Statistics, Abdul Wali Khan University Mardan, Pakistan, where he is currently pursuing the Ph.D. degree. His research interests include linear models, machine learning, applied statistics, and computational statistics.

HAFSA HINA received the M.Sc. degree in economics and finance from International Islamic University, Islamabad, in 2006, and the Ph.D. degree in economics from the Pakistan Institute of Development of Economics (PIDE), in 2013. She has worked as a Lecturer with the Department of Econometrics and Statistics, PIDE, from January 2010 to April 2014. She has been working as an Assistant Professor with the Department of Econometrics and Statistics, PIDE, since April 2014. Her research interests include time series econometrics, econometrics, applied econometrics, and microeconomics. She has been awarded the Gold Medal by International Islamic University, for her M.Sc. degree.

MUHAMMAD HAMRAZ received the bachelor's, master's, and M.Phil. degrees in statistics from Quaid-i-Azam University, Islamabad, Pakistan, in 2004, 2007, and 2010, respectively. He is currently an Assistant Professor of statistics with Abdul Wali Khan University Mardan, Pakistan. His research interests include linear models, biostatics, applied statistics, and computational statistics.

ZARDAD KHAN received the master's degree (Hons.) in statistics from the University of Peshawar, Pakistan, in 2008, the M.Phil. degree in statistics from Quaid-i-Azam University, Islamabad, Pakistan, in 2011, and the Ph.D. degree in statistics from the University of Essex, U.K., in 2015. He is currently an Associate Professor of statistics with Abdul Wali Khan University Mardan, Pakistan. He has also done a one year postdoctorate from the University of Essex. His research interests include machine learning, applied statistics, computational statistics, biostatistics, graphical modeling, and survival analysis.

SAEED ALDAHMANI received the bachelor's degree in statistics from United Arab Emirates University, United Arab Emirates, in 2007, the master's degree in statistics from Macquarie University, Australia, in 2010, the master's degree in applied finance from Western Sydney University, Australia, in 2011, and the Ph.D. degree in statistics from the University of Essex, U.K., in 2017. He is currently an Assistant Professor of statistics with United Arab Emirates University. His research interests include graphical models, biostatistics, applied statistics in finance, and computational statistics.