Machine Learning for Causal Inference and Environmental Policy Evaluation: An Application to Hedonic Property Analysis

Feryel Lassoued

***Abstract*— Water management policies can generate valuable services in the ecosystem but they are also costly to implement. Thus, we investigate this matter through the adoption of an irrigation water storage infrastructure in Chestermere lake. We adopt a hedonic property analysis and evaluate whether the implementation generates desirable attributes for nearby residential properties beyond irrigation services. Rooted in a causal machine learning approach, our hedonic property analysis captures the implicit value of the water quality by exploring the housing prices. To assess the causal impact of the policy, we construct a binary treatment variable and identify the waterfront properties in Chestermere that were sold after the implementation of the agreement. Following a hypothetical train-test-compare methodology, we employ an ensemble learning technique to predict the counterfactual selling price in the absence of the policy. Our method uses an extreme gradient boosting algorithm that conveys satisfactory performance in similar studies. We employ a grid search for reliable hyperparameters tuning and leverage a 10-fold cross-validation technique to evaluate our results. The derived learning algorithm yields a 91% accuracy rate that predicts an average treatment effect of $ATE = \$59,057.82$ with a 6.4% increase in the waterfront property values. Thereby, it acknowledges the positive and significant effect of the adoption of the water management agreement (WMA) in Chestermere beyond the irrigation services.**

***Index Terms*—ecosystem service, irrigation infrastructure, hedonic property analysis, water management policy, machine learning, non-market valuation, environmental policy analysis, causal inference**

I. Introduction

Non-market valuation for water resources is driven by the need for regulatory policy design or resource management decisions that might be the subject of litigation or environmental debates. Lakes and rivers generate valuable ecosystem services such as shoreline anchoring, scenic view and recreational activities. They provide household residents with substantial benefits. However, their quality and quantity differ based on the water resource's characteristics. The absence of a market governing these externalities makes the valuation of the water resource difficult. Subsequently, policy analysts refer to hedonic property analysis to provide a rationale for their investments.

Currently, in Alberta, Canada, Chestermere Lake depicts a prominent irrigation infrastructure. It originated as a water storage basin but has since fostered the growth of a town near the city of Calgary. The provision of the Lake aimed at enhancing and maintaining crop irrigation to provide food, fibre and income to the society. However, the fluctuation in the water levels and quality spur concerns among residents, Thereby leading to the implementation of a water management agreement (WMA) to preserve the lake's ecosystem services. The policy was carried out in September 2010 by an irrigation organization. Undoubtedly, these types of projects require expensive infrastructure given their scale. Albeit, they withhold benefits beyond irrigation such as desirable shoreline environments. Thus, we need to properly evaluate whether the benefits of the policy outweighs its cost.

Our research investigates the impact of water resource amenities on residential property values. We believe that the implementation of the WMA induces a positive effect over the shoreline housing prices given the generated upsurge in the environmental, aesthetic and recreational opportunities. We mainly investigate waterfront properties because of the lake's flat landscape that obscures the scenic view for non-shore estates. To capture the causal effect of the WMA relative to the monetary increase in the housing prices, we employ a hedonic analysis. We attempt to depict the trade-off between the lake's environmental services and the private real estate commodities. The monetary estimate generated from our study provides a proxy for welfare measures derived from a change in the quality of water ecosystem services. However, traditional hedonic approaches are plagued with omitted variable bias. Thus, to mitigate the bias, scholars usually refer to instrumental variables, fixed effects, or quasi-randomness method notwithstanding the underlying strong assumptions.

Meanwhile, some supervised machine learning (ML) algorithms are robust against multicollinearity and irrelevant variables. They can sustain high-dimensional data sets and can conduct systematic feature selection. They convey richer models with higher accuracy rates. Although traditional approaches primarily focus on prediction problems, a recent strand of the supervised machine learning literature is shifting towards causal inference.

Embedded in the fundamental problem of causal inference, ie no unit is observed in multiple counterfactual worlds at the same time, we employ a hypothetical train-test-compare methodology to discern the impact of the irrigation infrastructure management system. We exploit the non-parametric characteristics of decision trees and boosting technique to mitigate the bias in our estimate. In doing so, we

leverage an extreme gradient boosting algorithm that yields an accuracy rate of 91% and provides an empirical estimate of a 6.4% increase in monetary value.

Our methodology for valuing the adoption of the irrigation infrastructure agreement from the causal machine learning perspective is novel and will guide future research endeavours. It contributes to the literature of water resource policy analysis and the non-market policy valuation research.

II. Literature review

The Rosen hedonic model approach has been widely explored in non-market valuation schemes. Through regression analysis, the method quantifies ecosystem services that directly influence the housing market. Relevant literature investigating the relationship between property values and environmental goods mainly examines water quality attributes[2], [3], cancer clusters [4], fracking wells [5], crime [6], air quality [7], and power plants [8], While only a few measure the effect of lake levels. Landsford and Jones [9] were the first to explore the value of water in recreational and aesthetic uses through property prices around a lake in Texas. In hedonic studies, the property's proximity to the water resource is buttressed and recognised as a prominent feature. Given their numerous benefits to homeowners, housing prices will decrease as one moves away from the water resource [10], [9]. Loomis and fieldman [11] forecast the economic losses endured by lakeshore residents and the devaluation of their housing properties due to water level reduction. Likewise, Muller [12] estimates the premium price effect associated with lakes exhibiting constant water levels as opposed to the reduced monetary influence of lakes with periodic flooding, poor access, offensive odours and low water levels. The comparison between the two delineates the effect of scenic views and having access to nearby recreational activities on shoreline properties values.

Common hedonic property models, aiming to control for omitted variable bias, span from Quasi-experimental methods such as difference in difference [13], [5] and regression discontinuity [14] to equilibrium sorting models. [15]. However, to the best of our knowledge, relevant literature in the field neglects to investigate the use of supervised machine learning approaches notwithstanding their inherent resemblance to traditional statistical regression models. Methods of interest within the scope of our research vary from neural networks, regression trees, random forests, gradient boosting to regularized regression utilizing Lasso, Ridge or Elastic Net approaches. Ensemble methods combining the use of multiple learning algorithms gain momentum as together they outperform the single method approach in terms of their accuracy and predictiveness [16]. However, given the causal inference depicting the scope of our research, we explore these models beyond their predictiveness to be able to derive statistical inference. Recent studies within the field have emerged and are investigating causal relations. Chipman et al. [17] estimate, under the assumption of unconfoundedness, the impact of a binary treatment variable D on an outcome Y by controlling for the other characteristics X. In doing so, traditional statistical models produce biased estimates of the average treatment effect due to the presence of multicollinearity, irrelevant and/or omitted variable nuisances. In contrast, supervised machine learning research buttresses the automation of functional form and features selection.

Prominent approaches for estimating treatment effects are based on outcome regression and generalized propensity score models. Chipman et al. [17] and Athey et al.[18] proposed propensity score matching (PSM) methods for estimating causal effects, whereby they introduce a new feature denoted as the 'clever co-variate' C. PSM reduces the effect of treatment selection bias in observational studies. It investigates the subject's conditional probability of receiving a treatment effect rather than being assigned to a control group given a set of feature characteristics. To this end, machine learning algorithms were exploited and a double robust two-staged *Targeted Maximum Likelihood Estimator (TMLE)* was derived. The model generates an empirical density function of the outcome variable that is used to compute the average treatment effect. Athey [19] further proposed the use of Generalized Random Forests to compute non-parametric quantile regression, conditional average partial effect estimation, and heterogeneous treatment effect. *Bayesian Additive Regression Trees (BART)* also gain momentum as Generalized Random Forests [19] . Chernozhukov et al.[20] combine a regularization prior with regression trees and derive a predictive model that recursively partitions the features space into smaller subsets. In doing so, the model estimates $E(Y|D, X)$, computes the residuals and fits additional trees accordingly. Random forests approaches have also been investigated either directly [21],[22] or indirectly [23]. Varian et al. [24] highlight the extension of BART to *Bayesian Causal Forests (CRF)* whereby the authors improve the original model in terms of its induced regularization prior bias. Their BART model uses the propensity score estimate as a feature and introduces a function computing the treatment effect directly. It's noteworthy that both BART and BCF generate posterior distributions. Hahn[21] and Abadie[25] also used the generalized random forests to compute estimates of the treatment effect. The *Generalized random forest (GRF)* exploits the sample's average treatment effect $Y(D = 1)Y(D = 0)$ in the end node rather than the traditional outcome variable$Y$. GRF is robust in extracting the subsets of the variables and the values where the treatment effects would change the most and its derived estimates from averaging multiple trees are asymptotically normal. Likewise, Kim et al.[26] propose a *double /Debiased Machine Learning* method for causal and treatments effects. The double-robust approach incorporates a propensity score and an outcome regression, by expanding upon on the Frisch-Waugh-Lovell theorem. The model first predicts the outcome Y and the treatments D given the covariates X and then regress the residuals from $(E(Y|X))$ to the residuals induced by $(E(D|X))$. However, to prevent asymptotic bias in the residual-on-residual approach, the authors highlight the need for a "k-fold

cross- validation fitting. Furthermore, the *hypothetical train-test-treat-compare (TTTC) process* also gains momentum as scholars argue that causal inference, mainly, the counterfactual approach, is best suited for dealing with impediments to the effective use of machine learning techniques. Estimating the counterfactual is merely a prediction of what would have been in the absence of the treatment [27] [28]. The approach delineates the expected treatment effect as the difference between the counterfactual and the observed outcome.

III. Algorithm Description

Given the scope of our project and following the causal inference literature, we implement [29]'s hypothetical train-test-compare (TTTC) approach to assess whether the WMA policy induces monetary increments in housing prices. According to [29], the basic identity of causal inference is revealed through the difference in outcomes for the treated minus the controlled. We construct a binary treatment variable D that captures the waterfront properties in Chestermere which were sold after the implementation of the policy (WMA). This assignment separates the controlled observations from the treated samples and allows us to fit a predictive model to the controlled data set. This, in turn, provides an estimate of the counterfactual. Thereafter, we test the model's predictiveness power and assess how well it performs out of the sample. Eventually, taking the difference between the predicted counterfactual and the observed prices for the treated properties yields an estimate for the average treatment effect of the policy. To do so, we leverage gradient boosting regression trees.

A. Regression trees

Regression trees are one specific form of decision trees that automatically investigate nonlinearities and interactions among the covariates. They systematically partition the feature space into separate regions with similar observations. This process is known as binary recursive partitioning, whereby the algorithm assigns data to the two primary partitions with all possible binary splits for every feature. With each split, the average outcome of the partitioned regions is computed and assigned as the predicted value. Regression trees can handle high-dimensional datasets at no cost to the model and their choice for feature selection is governed by their ability to predict the outcome. It predicts very well out-of-sample as they average many trees together. However, averaging across a substantial number of trees increases prediction accuracy at the cost of interpretability. Subsequently, they can easily sustain multicollinearity issues. In each split, the model aims to minimize the sum of squared residuals within the region it creates. Given an outcome y_i, a tree is depict using observations $x_i = (x_{i1}, x_{i2}, ..., x_{ip})$ for $i = 1, 2, .., N$ and p features :

$$f(x) = \sum_{j=1}^{J} \gamma_j I(x \in R_j)$$

where $f(x)$ conveys the predicted outcome. The feature space is partitioned into J regions(R_j). I is an indicator function for whether x pertains to the split region R_j and γ_j is chosen to minimize the Sum of Squared Residuals (SSR), $\sum(y_i - f(x_i))^2$, whereby the average outcome for all samples within the region R_j constitutes the optimal choice for γ_j. The split of the feature space is driven by optimal cutoffs and is carried out through a single feature at a chosen threshold. With each decision boundary, the regression tree algorithm investigates all features and thresholds, and then selects the variable x_k and threshold s that minimize the sum of the squared residuals in the two regions $R_1(k, s)$ and $R_2(k, s)$

$$R_1(k, s) = \{x | x_k < s\} \text{ and } R_2(k, s) = \{x | x_k > s\}$$

Splits are carried out sequentially on the regions and their descendant decision boundaries. These are greedy algorithms and once a split has been executed, no alterations to previous boundaries can be introduced. Nevertheless, numerous splits for the same feature at distinct thresholds can be derived. The chosen limit on the minimum number of leaves (or terminal nodes or final regions) and/or the number of remaining data points within each region convey the model's tuning parameters. Overall, maintaining fewer data points within each region and constructing more terminal nodes yields a lower bias and a higher accuracy for the training set. However, predictions across these tree splits are very sensitive to the training data and subsequently tend to have high variance. Hence, the need for ensemble learning methods.

B. Ensemble Learning

In ensemble learning theory, weak learners (or base models) constitute the building blocks of more complex models. Combining and training several weak learners using the same learning algorithm generates a more robust and stronger learner (or ensemble model). High bias or high variance are mainly incurred when considering only a single decision tree. Therefore, the ensemble model aims to reduce the bias and/or variance of the weak learners by combining them together. This results in better predictive performance and accuracy. The bias-variance trade-off is driven by the need for having enough degrees of freedom to investigate the underlying complexity of the data while maintaining low variance. (Fig 1)

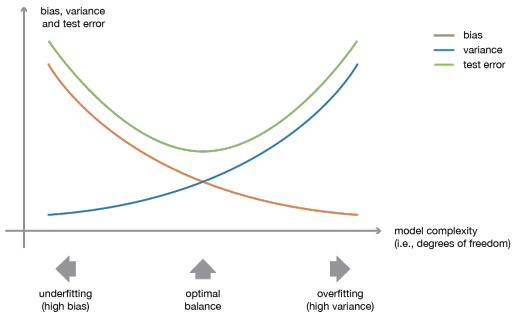

Fig. 1. The bias-variance trade off

The choice of weak learners must be coherent with the aggregating method. Low bias and high variance base models should be combined in a way that renders the ensemble model more robust whereas low variance and high bias weak learners should be consolidated in a way that generates a less biased aggregate model. Relevant meta-algorithms range from bootstrap aggregation (Bagging), boosting, cascading models to stacked Ensemble Models. However, given our weak learner choice, we will mainly investigate boosting techniques.

C. *Gradient Boosting Machines*

In sequential methods, boosting generates an ensemble model that is, in general, less biased than the weak learners that composes it. The consolidated weak learners are no longer fitted independently from each other rather sequentially in a very adaptive way. The training of the model at a given step depends on the previously fitted model, whereby higher weights are given to observations that were poorly fitted by previous models in the sequence. Essentially, each new model focuses on the most challenging data points in order to derive a stronger learner with lower bias. However, boosting does not support parallel computing which renders sequentially fitting manifold complex models computationally expensive and makes reaching the optimal model too difficult. Hence, the need for an iterative approach.

Friedman (2001) introduces gradient boosting as an additive ensemble method, whereby a weighted sum of weak learners is added together. Grounded in a gradient descent approach, gradient boosting easily adapts to various loss functions. With decision tree models used as base learners, the procedure appends a new tree to the weighted sum of all previous trees and updates observations at each iteration. The first iteration is initialized with the average of the outcomes $\bar{y}$ for N observations and with each new iteration, weak learners are trained to fit the pseudo-residuals that guide the direction and correct the current ensemble model predictions in order to lower the error. Given M iterations, a tree $f_m(x)$ and a shrinkage parameter ν aiming at slowing down the learning rate, preventing the model from overfitting, and hindering any iteration from gaining too much influence, we derive the following predictive model:

$$F_M(x) = \sum_{m=1}^{M} \nu f_m(x)$$

A small ν is between the 0.01 and 0.1 thresholds is usually chosen in order to assign a low weight on each additional tree. It generates a less sensitive prediction. The optimal model is driven by the choice of M and ν with higher values of M requiring lower ν. Likewise, sub-sampling at each iteration mitigates overfitting and reduces the variance of the predicted outcome. The tree $f_m(x)$ omits one feature from the input space and is more of a stump. The depth of these trees can be increased, however, they perform best with fewer terminal nodes.

The base learner regression tree $h_m(x)$ is used to fit the residuals $r_m(x)$, with the step size or weight γ_m chosen to minimize the loss at iteration m. Assuming an easily differentiable sum of squared residuals loss function $L(y, F(x)) = -\frac{1}{2}(y - F(x))^2$ and an $f_m(x) = -\gamma_m h_m(x)$, our ensemble learning model is as follows:

$$F_m(x) = F_{m-1}(x) + \nu \quad \gamma_m \quad h_m(x)$$

Table 1 further delineates the step by step algorithm.

TABLE I
GRADIENT BOOSTING ALGORITHM

Inputs:

- Training set $\{(x_i, y_i)\}_{i=1}^{n}$
- Differential loss function L(y,F(x))
- Number of iterations M

Algorithm:

1) Initialize the model with a constant value

$$F_0(x) = argmin_\gamma \; L(y_i, \gamma)$$

2) For $m = 1$ to M do:
   a) Compute the pseudo-residuals

$$r_{im} = y - F_m(x) = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x)}\right]_{F=F_{m-1}} \quad for \; i = 1, ., N$$

   b) Fit a regression tree (the base learner) $h_m(x)$ to the pseudo residuals using the training set $\{(x_i, y_i)\}_{i=1}^{n}$
   c) Compute the multiplier by solving the following one-dimensional optimization problem :

$$\gamma_m = argmin_\gamma \sum_{i=1}^{N} L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i))$$

   d) Update the model:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)$$

3) Output the predicted outcome for observation x: $F_M(x)$

Extreme Gradient Boosting is a specific implementation of Gradient Boosting. It investigates more accurate estimates in order to derive the best tree model. By computing second-order gradients of the chosen loss function, the model learns the direction of gradients and minimizes its error The learning is either applied through Ridge or Lasso regularization to improve model generalization, address over-fitting and guide feature selection .

While machine learning approaches are prone to overfitting to the training samples, analysing the mean absolute error (MAE) for an out-of-sample test set allows us to assess the accuracy of our model. To this end, cross-validation is carried out to ensure better accuracy and refrain from peeking.

IV. DATA DESCRIPTION

We were provided with transaction Real Estate data for the Town of Chestermere and the city of Calgary. The Calgary time series data set covers observations from 1981 to 2010. However, concerns regarding the extent of the market lead us to omit samples from 1981 till 1999. Each property is described by its structural characteristics namely its

residential area, lot size, number of bedrooms, bathrooms and geographical coordinates and the distance from the property to the Lake. Waterfront characteristics for these properties along with the presence of some other housing characteristics are covered as well. The variable sales price has been adjusted for inflationary effects for 2007 dollars using Calgary's new housing price index (NHPI) which was provided by Statistics Canada. Categorical dummy variables are adopted to describe the house age, garage spaces, and the presence of fireplaces. The water management agreement policy was implemented in September 2005. Subsequently, the constructed binary treatment variable D identifies the waterfront properties in Chestermere that were sold after the implementation of the policy (WMA), which in turn allows us to assign our control and treatment groups. Upon treating outlier observations due to incorrect structural and geographical information, we obtain a data set of 219,131 sample sales, whereby 219,084 data points constitute the control group and the remainder 47 capture the treated observations.

Table *II* delineates the definition of these variables and provides the mean sample values for the control and treated groups.

TABLE II
DESCRIPTIVE STATISTICS OF THE CONTROL AND TREATMENT GROUPS

Name of the variable	Description of the variable	Mean Values	
		Treatment Group	Control Group
sold price	House sales prices (2007 $CAD)	922,512.77	303,915.74
sold year	The year when the property was sold	2007.28	2005.20
sold month	The month when the property was sold	6.77	6.21
yrbuilt	The year when the property was built	1988.34	1984.57
lat	Latitude coordinates of the property	51.04	51.04
long	Longitude coordinates of the property	-113.82	-114.08
*group**	1 for Chestermere, 0 for Calgary	1.000000	0.008
area	Square meters of area not including basement	218.28	127.42
bedroom	Number of bedrooms	3.68	3.09
bathroom	Number of bathrooms	2.74	1.92
*ac**	Presence of air conditioning system	0.21	0.06
*decbal**	Presence of Deck or balcony	0.85	0.61
*waterf**	Waterfront property	1	0.003
*sfhouse**	Single family houses	0.94	0.71
*summer**	Houses that are sold in June through August	0.298	0.27
houseage	Houses age	18.94	20.64
*dehouse**	House types: detached or attached	0.94	0.64
mtolake	Distance (meters) from each house to the lake	103.096	19010.981
szg	No. of cars to be stored in the garage	2.17	1.11
fireplace	No. of fireplaces	1.40	0.64
*wma**	Water management agreement:(Sep 2005)	1	0.51
*D**	Treatment variable	1	0
N	No. of observation	47	219,084

*Binary variables

The relationship between the town of Chestermere property locations and their selling price is depicted in figure 2. The analysis concludes that the highest valued properties in the town of Chestermere are shoreline properties that were sold after the implementation of the policy.

Fig. 2. Relationship between sales prices of properties in the Town of Chestermere and their distance to Chestermere Lake

Figure 3 also delineates the increase in shoreline property values after the implementation of the policy.

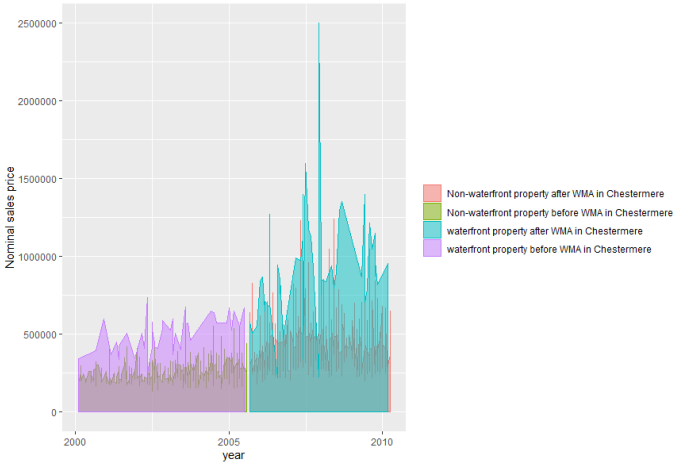

Fig. 3. Price (2007 $CDN) of properties sold in the Town of Chestermere from 2000 to 2010

Moreover, a closer look into the selling price distribution for the control and treatment group reveals that while most of the treated properties were sold for less than 2 million, the controlled houses were marketed below 1 million CAD. Selling prices are skewed to the right and some outliers lie above 3 million. Fig(4)

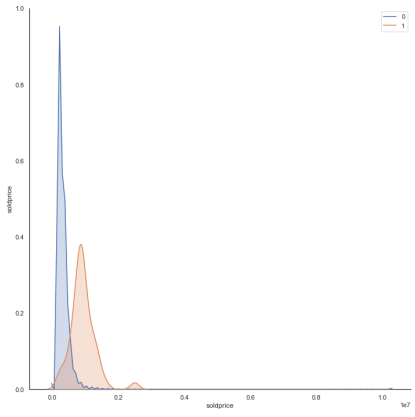

Fig. 4. Selling price distribution for the control and treatment group

Furthermore, we derive a heat map depicting the underlying relationships among our explanatory features and our dependant variable 'sold price'. We deduce that the selling price is mostly correlated with the area, the number of bathrooms, garage size and the number of fireplaces in the house. (Fig 5)

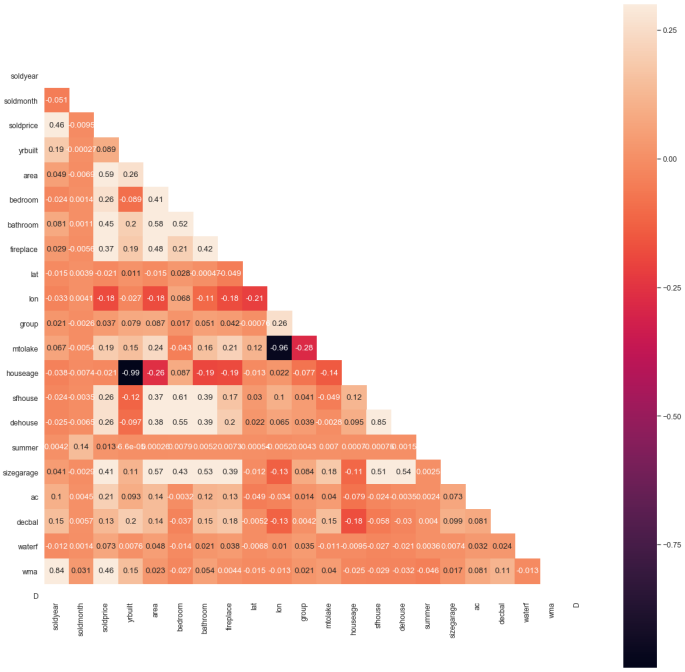

Fig. 5. Heat-map of Correlations amongst the variables in the control group

The following scatter plots further describe the underlying relationship relating property characteristics to their selling prices for the treated and controlled groups. A relatively high proportion of properties had a zero value for house age because many houses were newly built. We conclude that there is no clear linear pattern that would allow us to explain the variability in the selling price. (Fig 6,7)

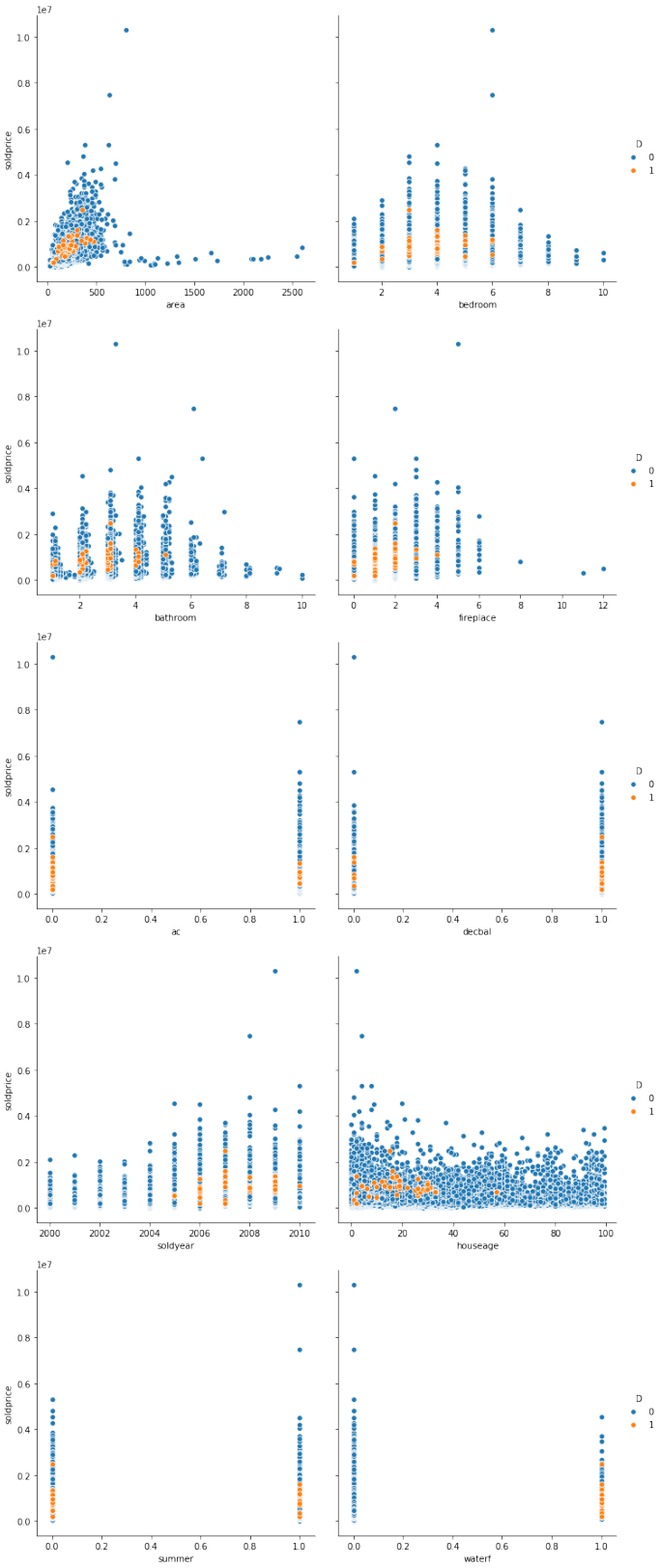

Fig. 6. Relationship between sales prices and house characteristics for the control and treated properties

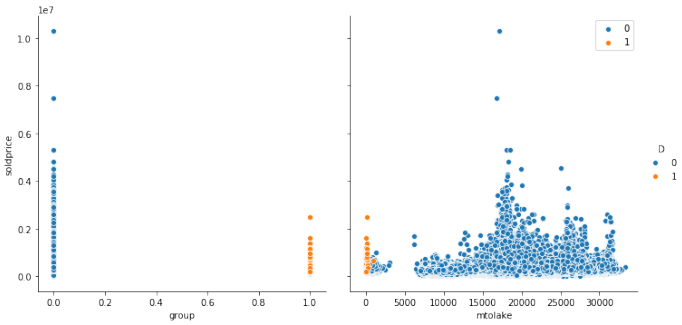

Fig. 7. Relationship between sales prices and house characteristics for the control and treated properties

V. Results

Our XGBoost learning algorithm was run with a 1000 tree estimators and a learning rate of 0.1. We applied an L2 regularization of 0.8 to prevent from overfitting. The algorithm splits up to a maximum depth of 6 nodes in each tree. Albeit, these values were not obtained arbitrarily rather through the process of a grid search. Table II provides a summary of the inferred hyperparameters from the grid search approach.

TABLE III
Grid search results for hyperparameter tuning

Hyperparameter	Candidates	Best HP
Colsample_Bytree	$0.6, 0.8, 1$	1
Gamma	$0, 0.03, 0.1$	0
Max_Depth	$1, 1.5, 6$	6
Learning_Rate	$0.1, 0.07$	0.1
Min_Child_Weight	$5, 6, 8$	8
N_Estimators	$1000, 1500, 2000$	1000
Reg_Alpha	$0, 1e^{-5}, 1e^{-2},$	0
Reg_Lambda	$1e^{-2}, 0.5, 0.8$	0.8
Subsample	$0.6, 0.95, 1$	0.95

Hyperparameters are crucial for the model's performance. They convey the algorithm's characteristics that cannot be estimated from the data directly, rather they must be chosen prior to the learning process. To infer the optimal values, we leverage a grid search and derive better accuracy results. While the grid search iterates over all possible combinations, a model is built per the input candidate parameters. Our chosen candidate parameters stem from the literature and a trial & error approach. Throughout the process, we implement a 10-fold cross-validation to assess the choice of the hyperparameters and the robustness of the model in terms of prediction performance. It also helps to concrete the estimation of parameters imparting any kind of bias or overfitting.

Upon leveraging the optimum values from the grid search, we obtain a training mean absolute error of $26,736.85$ which is consistent with our $26,191.8$ out of sample MAE loss. The learning algorithm yields a 91% accuracy rate.

Figure 8 depicts the first boosted decision tree in our ensemble model within the learning sequence, whereby the feature space values governing split decisions in each node and output leaves are inferred. The variables are systematically identified by f_i per the feature index in the input array.

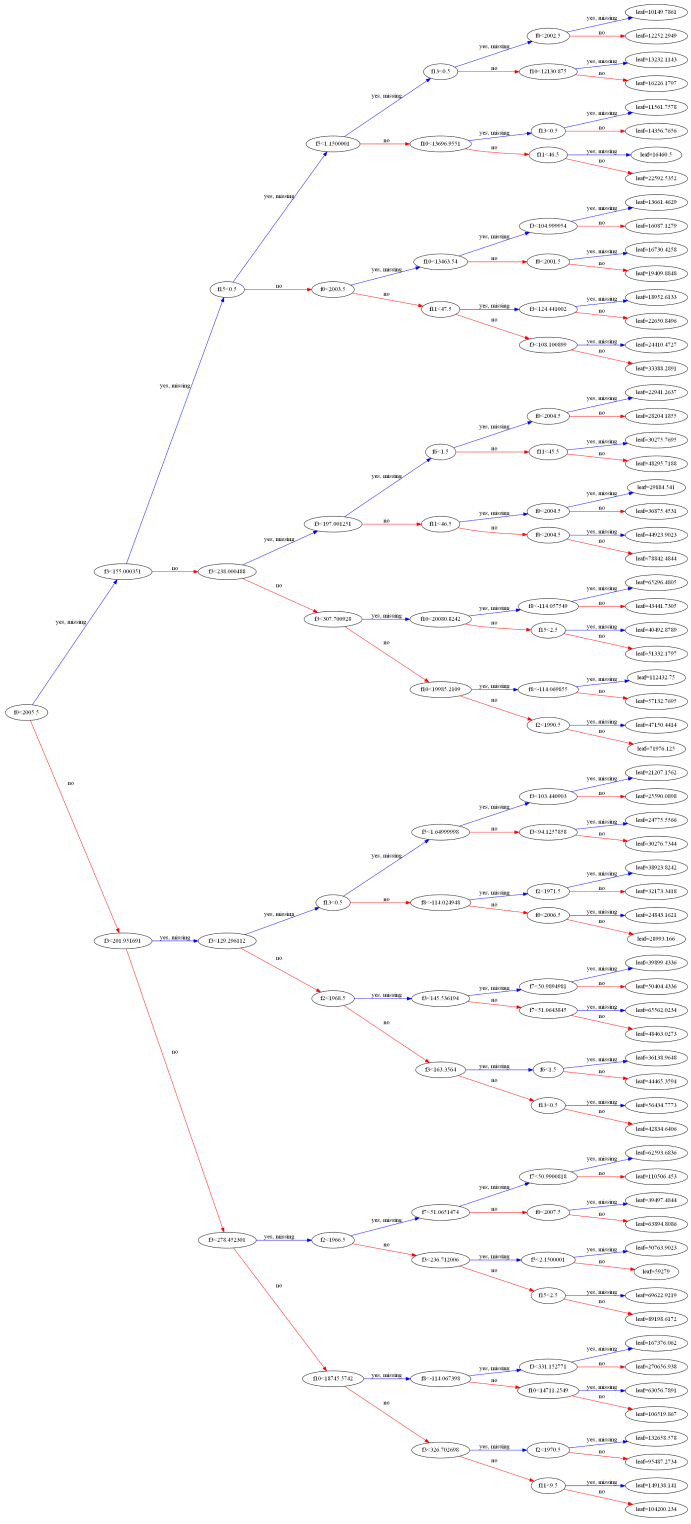

Fig. 8. XGBoost Plot of Single Decision Tree

VI. Discussion

Per Varian's [24] hypothetical train-test-compare (TTTC) approach, we compute the average treatment effect on the treated property as follows:

$$ATE = \frac{1}{N^{Treat}} \sum_{1}^{N^{Treat}} (Y^{Treat} - \hat{F}(x))$$

$\hat{F}(x)$ predicts the counterfactual and delineates the estimated selling price of the treated properties, $\hat{Y}^{Treat}$refers to the realized observation and N^{Treat} denotes the number of waterfront properties that were sold after the implementation of the water management agreement in the town of Chestermere. Subsequently, the average treatment effect is computed from the average residuals of the treated houses. Tree models match the treatment group's data samples to the control group and are assigned to regions per their feature values. Given the average outcome of the observations in that region, the counterfactual is estimated through boosting and weighting the sum of relevant trees. Nevertheless, the treated samples are not matched to just one data point in the control group but to a group of observations. This procedure is consistent with [25]'s synthetic control group design, whereby the authors estimate the counterfactuals by matching to a group of control samples per a set of matching features.

Our TTTC approach for deriving causal inference with respect to environmental policy valuation yields an average treatment effect estimate of $ATE = \$59,057.82$. The results provide evidence and buttress our hypothesis that the implementation of the WMA in Chestermere induced a real and significant effect on the values of the shoreline housing prices. The percentage increase in property values for the lake is estimated to be 6.4% on average.

Figure 9 further illustrates the causal impact of the policy as it contrasts the counterfactual to the actual observed market prices for the treated properties.

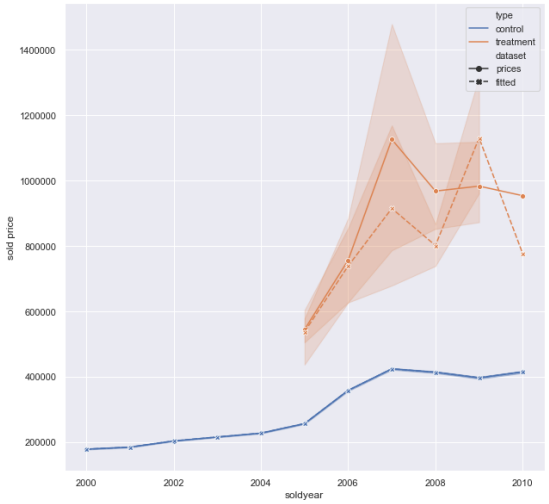

Fig. 9. Treated and control groups counterfactual and property prices

Across the years, the fitted values predict the real estate market with a 91% accuracy rate and manage to capture relevant trends. The selling price for Chestermere waterfront houses trends upwards after the signing of the WMA in 2005. However, prices of non-waterfront properties incur a relatively smaller increment in comparison. The results also delineate a lag in the treatment effect, whereby 4 months after the adoption of the policy, we start to discern the gradual deviation of the counterfactual from the observed. Subsequently, the results of the study confirm that the implementation of the WMA generated increased tax revenue for the town, thereby allowing the settlement to internalize the positive externalities of their investments.

Plotting the covariates' contribution and relative relevance in deriving the model's prediction further buttress this rationale (Fig 10)

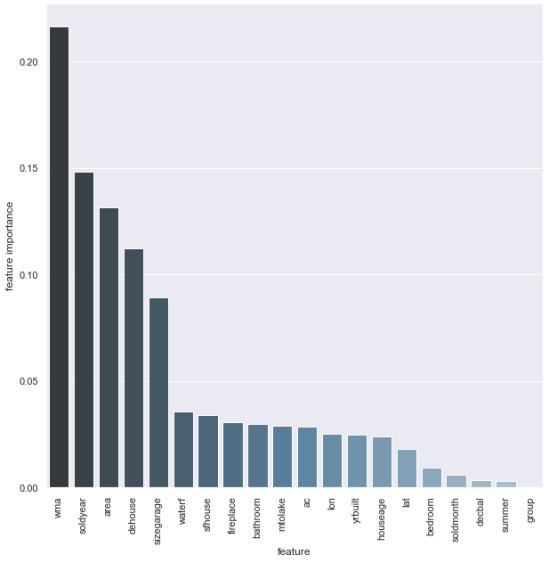

Fig. 10. Feature weights

The signing of WMA dominates the other features and stands out as the most important predictor of our housing price market. These findings buttress the impact of the policy on all properties in the dataset. However, the difference in the upward shift between the treatment and control group delineates the existence of a direct and indirect effect associated with the policy. Grounded in the assumption that only the town's shoreline properties could benefit from the view, our research mainly investigated the average treated effect on the 47 waterfronts Chestermere houses that were sold after the adoption of the policy. To this end, we estimated an average treatment effect of $AVT = \$59,057.82$ and a 6.4% increase in property values. On the other hand, the original Diff and Diff study [26]sought out to capture the average treatment effect on all 389 waterfront properties in the dataset. Their study yielded an average estimate of $\$48,299$ coupled with a 5.5% increase rate.

VII. Conclusions and future work

This research investigates the differential impact of a water management agreement using the case of irrigation infrastructure at Chestermere Lake. In doing so, we mainly focus on distinguishing the primary effect of the stabilization of water levels and quality on the sales prices for the surrounding shoreline Chestermere properties. To this end, we derive evidence that the ecosystem service generated by the irrigation infrastructure provides a substantial positive economic impact on the chosen properties.

Notwithstanding the small number of properties directly affected by the water management policy, the hypothetical train-test-treat-compare (TTTC) approach was able to discern the average treatment effect for the 47 chosen observations with a 91% accuracy rate. In doing so, our methodology integrates the econometrics of hedonic property analysis, machine learning and the science of causal inference. Machine learning algorithms can easily sustain confoundedness, omitted variables bias, feature selection and multicollinearity issue. They allow us to treat outliers without the need for any strong functional form assumption and they can automatically explore nonlinearities and interactions across variables for effective model selection. Subsequently, machine learning for causal inference and environmental policy valuation should be further explored within the literature, given the growth in the availability of data in the form of text, images and voices.

Future work entails the possible use of housing images to expand beyond the classical features and control for more attributes. A richer model would be derived to better sustain the omitted variable bias in hedonic property analysis. We could also leverage coordinate descent and genetic algorithms to derive better hyperparameters and higher model accuracy in and out of the sample. Likewise, benchmarking multiple approaches such as Targeted Maximum Likelihood Estimator, Generalised random forest, Double machine learning and Bayesian Causal Forests would generate a more robust analysis.

VIII. References

[1] S. Athey, "Machine Learning and Causal Inference for Policy Evaluation," in Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '15, Sydney, NSW, Australia, 2015, pp. 5–6, doi: 10.1145/2783258.2785466.

[2] K. J. Boyle, P. J. Poor, and L. O. Taylor, "Estimating the Demand for Protecting Freshwater Lakes from Eutrophication," Am. J. Agric. Econ., vol. 81, no. 5, pp. 1118–1122, Dec. 1999, doi: 10.2307/1244094.

[3] C. G. Leggett and N. E. Bockstael, "Evidence of the Effects of Water Quality on Residential Land Prices," J. Environ. Econ. Manag., vol. 39, no. 2, pp. 121–144, Mar. 2000, doi: 10.1006/jeem.1999.1096.

[4] L. W. Davis, "The Effect of Health Risk on Housing Values: Evidence from a Cancer Cluster," Am. Econ. Rev., vol. 94, no. 5, pp. 1693–1704, Nov. 2004, doi: 10.1257/0002828043052358.

[5] L. Muehlenbachs, E. Spiller, and C. Timmins, "The Housing Market Impacts of Shale Gas Development," Am. Econ. Rev., vol. 105, no. 12, pp. 3633–3659, Dec. 2015, doi: 10.1257/aer.20140079.

[6] L. Linden and J. E. Rockoff, "Estimates of the Impact of Crime Risk on Property Values from Megan's Laws," Am. Econ. Rev., vol. 98, no. 3, pp. 1103–1127, May 2008, doi: 10.1257/aer.98.3.1103.

[7] K. Y. Chay and M. Greenstone, "Does Air Quality Matter? Evidence from the Housing Market," J. Polit. Econ., vol. 113, no. 2, pp. 376–424, Apr. 2005, doi: 10.1086/427462.

[8] L. W. Davis, "The Effect of Power Plants on Local Housing Values and Rents," Rev. Econ. Stat., vol. 93, no. 4, pp. 1391–1402, Nov. 2011, doi: 10.1162/REST$_{a0}$0119.

[9] N. H. L. Jr and L. L. Jones, "Recreational and Aesthetic Value of Water Using Hedonic Price Analysis," p. 16.

[10] B. L. Mahan, S. Polasky, and R. M. Adams, "Valuing Urban Wetlands: A Property Price Approach," Land Econ., vol. 76, no. 1, p. 100, Feb. 2000, doi: 10.2307/3147260.

[11] J. Loomis and M. Feldman, "Estimating the benefits of maintaining adequate lake levels to homeowners using the hedonic property method: ECONOMIC BENEFITS OF LAKE LEVELS," Water Resour. Res., vol. 39, no. 9, Sep. 2003, doi: 10.1029/2002WR001799.

[12] N. Z. Muller, "Using hedonic property models to value public water bodies: An analysis of specification issues: USING HEDONIC PROPERTY MODELS TO VALUE PUBLIC WATER BODIES," Water Resour. Res., vol. 45, no. 1, Jan. 2009, doi: 10.1029/2008WR007281.

[13] Y. Kim and P. Steiner, "Quasi-Experimental Designs for Causal Inference," Educ. Psychol., vol. 51, no. 3–4, pp. 395–405, Oct. 2016, doi: 10.1080/00461520.2016.1207177.

[14] S. E. Black, "Do Better Schools Matter? Parental Valuation of Elementary Education," Q. J. Econ., vol. 114, no. 2, pp. 577–599, May 1999, doi: 10.1162/003355399556070.

[15] J. M. Duke, J. Wu, H. A. Klaiber, and N. V. Kuminoff, "Equilibrium Sorting Models of Land Use and Residential Choice," in The Oxford Handbook of Land Economics, J. M. Duke and J. Wu, Eds. Oxford University Press, 2014.

[16] J. L. Hill, "Bayesian Nonparametric Modeling for Causal Inference," J. Comput. Graph. Stat., vol. 20, no. 1, pp.

217–240, Jan. 2011, doi: 10.1198/jcgs.2010.08162.

[17] H. A. Chipman, E. I. George, and R. E. McCulloch, “BART: Bayesian additive regression trees,” Ann. Appl. Stat., vol. 4, no. 1, pp. 266–298, Mar. 2010, doi: 10.1214/09-AOAS285.

[18] S. Athey and G. W. Imbens, “Machine Learning Methods That Economists Should Know About,” Annu. Rev. Econ., vol. 11, no. 1, pp. 685–725, Aug. 2019, doi: 10.1146/annurev-economics-080217-053433.

[19] S. Athey, J. Tibshirani, and S. Wager, “Generalized Random Forests,” ArXiv161001271 Econ Stat, Apr. 2018, Accessed: Apr. 08, 2020. [Online]. Available: http://arxiv.org/abs/1610.01271.

[20] V. Chernozhukov, D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, and W. Newey, “Double/Debiased/Neyman Machine Learning of Treatment Effects,” ArXiv170108687 Stat, Jan. 2017, Accessed: Apr. 08, 2020. [Online]. Available: http://arxiv.org/abs/1701.08687.

[21] P. R. Hahn, C. M. Carvalho, J. He, and D. Puelz, “Regularization and confounding in linear regression for treatment effect estimation,” ArXiv160202176 Stat, Dec. 2016, Accessed: Apr. 08, 2020. [Online]. Available: http://arxiv.org/abs/1602.02176.

[22] S. Athey, J. Tibshirani, and S. Wager, “Generalized random forests,” Ann. Stat., vol. 47, no. 2, pp. 1148–1178, Apr. 2019, doi: 10.1214/18-AOS1709.

[23] V. Chernozhukov, D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, and W. Newey, “Double/Debiased/Neyman Machine Learning of Treatment Effects,” Am. Econ. Rev., vol. 107, no. 5, pp. 261–265, May 2017, doi: 10.1257/aer.p20171038.

[24] H. R. Varian, “Big Data: New Tricks for Econometrics,” J. Econ. Perspect., vol. 28, no. 2, pp. 3–28, May 2014, doi: 10.1257/jep.28.2.3.

[25] Abadie, A., Gardeazabal, J. (2003). The economic costs of conflict: A case study of the Basque Country. American economic review, 93(1), 113-132.

[26] No Kim, Hyun, Peter C. Boxall, and W. L. Adamowicz. ”The demonstration and capture of the value of an ecosystem service: A quasi-experimental hedonic property analysis.” American Journal of Agricultural Economics 98.3 (2016): 819-837.