Heterogeneous Effects of US Medicaid Expansions on Health Insurance Coverage

Victor Wang*
University of Cambridge

We use cross-sectional data from the 2014 American Community Survey to estimate the heterogeneous treatment effects of expanding Medicaid on health insurance coverage. In doing so, we provide robust evidence which can help policymakers target future Medicaid policies towards particularly responsive individuals. Medicaid expansions were optional by state, allowing us to identify treatment effects by comparing expansion and non-expansion states. We then estimate heterogeneous treatment effects using a non-parametric machine learning algorithm called a causal forest, which offers several advantages over prior methods in the literature. Most notably, it provides a systematic means to discover, from the data, which variables are most relevant for modelling heterogeneity. We find strong evidence of heterogeneity in our estimated treatment effects. Furthermore, we find that individuals with the largest treatment effects were typically aged 25-34; had a high school diploma as their highest level of education; spoke a language other than English at home; and/or were in private for-profit employment.

I. INTRODUCTION

As of 2020, 31.6 million individuals in the US do not have health insurance(Cha and Cohen, 2022). This comprises 9.7% of the US population(Cha and Cohen, 2022), which is substantially higher than uninsurance rates in other developed countries(Papanicolas, and Jha, 2018). Lack of health insurance is linked to negative outcomes including worse overall health and higher mortality rates(Coleman, 2006).

The US government has made extensive efforts to tackle this problem, particularly through the 2014 implementation of the Affordable Care Act (ACA). One key provision of the ACA aimed to significantly expand eligibility for Medicaid, a public health insurance programme, in order to provide affordable insurance to almost all adults under age 65 with incomes below 138% of the federal poverty line(Garthwaite, Graves, Gross, Karaca, Marone and Notowidigdo, 2019). However, these Medicaid expansions were made optional for states, and only 26 states actually implemented the expansions in 2014. This suggests an observational study using cross-sectional data from 2014 to compare expansion and non-expansion states, thereby allowing us to identify the treatment effect of Medicaid expansions on the proportion of individuals covered by health insurance [1].

We are further interested in estimating *heterogeneity* in this treatment effect. That is, we seek to model the treatment effect on individuals as a function of their covariates. By doing so, we can identify which individuals experienced the largest treatment effects in response to the Medicaid expansions. Such information would allow future Medicaid policies to be targeted towards particularly responsive individuals. This is especially relevant in settings where policymakers have limited financial resources, and therefore can only afford to expand Medicaid to a limited number of individuals.

A key methodological contribution of this paper is to estimate heterogeneous treatment effects using a machine learning technique called a causal forest. This method has not been applied to study the ACA Medicaid expansions before, and it offers important advantages over more traditional approaches in the prior literature.

Notably, causal forests provide a flexible and systematic way to discover, based on the data, which variables are most relevant for modelling heterogeneity. In more traditional methods, heterogeneity can only be estimated over a small number of pre-specified variables, meaning researchers might miss important variables that they did not specify at the outset of their study. Causal forests also allow for non-parametric estimation of the treatment effects, meaning researchers can be agnostic regarding functional form and let the algorithm determine this from the data.

We use a large dataset from the 2014 American Community Survey, which contains information on over 3 million Americans nationwide. In addition to documenting the insurance status of each surveyed individual, it records a rich set of individual socioeconomic, demographic and health-related covariates. We also include several state-level controls in our study to minimise any confounding.

We find strong evidence in favour of heterogeneity in our estimated treatment effects. Furthermore, we find that individuals with the largest treatment effects were typically aged 25-34; had a high school diploma as their highest level of education; spoke a language other than English at home; and/or were in private for-profit employment. This last variable was not generally noted in prior studies, and its discovery demonstrates how novel machine learning approaches can bring new evidence to the existing literature.

The remainder of the paper is structured as follows. Section II provides an overview of Medicaid and the ACA; it then discusses the policy relevance of estimating heterogeneous treatment effects, considers the ben-

* vwang57@gmail.com

[1] We subsequently refer to this outcome variable as 'insurance coverage' for brevity.

efits of a machine learning approach, and ends with a brief review of the related literature. Section III presents our methodology, including a detailed description of the causal forest algorithm. Section IV details our identification strategy and dataset. Section V presents our key results, and Section VI concludes.

II. BACKGROUND, MOTIVATION AND RELATED LITERATURE

A. Overview of medicaid and the affordable care act

Medicaid is a public health insurance programme in the US, which is jointly funded by states and the federal government. It covers healthcare costs for certain individuals with limited income or resources. Medicaid is administered by states, subject to federal requirements(Center for Medicaid and CHIP Services, 2022).

Historically, Medicaid-eligible individuals had to have low income and belong to a certain category – such as being a disabled individual, a parent, a pregnant woman, or a child(Gruber and Sommers, 2019). However, eligibility was substantially altered by the Affordable Care Act (ACA). This federal statute introduced a far-ranging package of healthcare reforms; it was signed into law in 2010 and mostly implemented in 2014. A key goal of the ACA was to maximise insurance coverage(Courtemanche, Marton, Ukert, Yelowitz and Zapata, 2017).

One major provision of the ACA was intended to largely eliminate categorical criteria for Medicaid eligibility(Garthwaite, Graves, Gross, Karaca, Marone and Notowidigdo, 2019), by extending the programme to include nearly all adults under age 65 with incomes below 138% of the federal poverty line. However, a Supreme Court ruling in 2012 made this Medicaid expansion optional for states(Courtemanche, Marton, Ukert, Yelowitz and Zapata, 2017). Indeed, 24 states opted out during the initial implementation of the ACA in 2014. States which did not expand Medicaid generally retained their historic eligibility criteria, including categorical requirements.

This institutional setting allows us to identify the treatment effect of Medicaid expansions on insurance coverage, by comparing outcomes in expansion and non-expansion states[2].

B. Policy insights from estimating heterogeneous treatment effects

In this study, we will estimate *heterogeneity* in the treatment effects of Medicaid expansions on insurance coverage. That is, we model the treatment effects on individuals as a function of their covariates, including factors such as age, gender, or race. This allows us to find which subgroups[3] of the population experienced the largest treatment effects in response to expanding Medicaid. Below, we discuss two scenarios where policymakers could use this information to target future Medicaid policies to particularly responsive subgroups.

Firstly, several states opted out of the ACA Medicaid expansions in 2014 on the grounds that such expansions would be too expensive to implement. Some of these states, such as Georgia, have subsequently expressed interest in much smaller-scale Medicaid expansions, targeting specific subsets of the population(Hellmann, 2019). By focusing on particularly responsive subgroups, policymakers can maximise coverage gains, while keeping the expansions small-scale enough to minimise financial cost.

Secondly, states may have to disenrol individuals from Medicaid after the end of the Public Health Emergency (PHE) that was first declared in 2020 due to Covid-19. Federal funding for Medicaid has been enhanced for the duration of the PHE only, conditional on a 'continuous coverage requirement': anyone enrolled in Medicaid after March 18 2020 cannot be disenrolled during the PHE(Alker and Brooks, 2022).

The PHE is currently anticipated to end around mid-to-late 2022. Subsequently, states might scale back their Medicaid programmes due to reduced federal funding. Furthermore, the continuous coverage requirement will no longer apply, so states will reassess whether individuals are still eligible for Medicaid(Corlette and Kona, 2022). For both reasons, states may have to disenrol individuals from Medicaid(Buettgens and Green, 2022). However, policymakers may wish to avoid disenrolling particularly responsive subgroups. Ensuring this will require robust evidence on treatment effect heterogeneity.

C. Estimating heterogeneous treatment effects with machine learning

We estimate heterogeneous treatment effects using machine learning, which offers key advantages over more traditional approaches. One such advantage is that machine learning algorithms can estimate treatment effects non-parametrically, allowing the researcher to be agnostic regarding functional form.

Another advantage is that machine learning algorithms permit the researcher to be agnostic pre-analysis about which variables are most relevant for modelling treatment effect heterogeneity. Instead, the algorithm will automat-

[2] Section IV A details our identification strategy.

[3] A subgroup is a subset of the population, typically defined with reference to individual covariate values. For instance, if age is an individual covariate, then one potential subgroup could be 'all individuals in the population aged below 30'.

ically discover such variables during the analysis phase, based on the data.

Without the use of machine learning, analysing subgroups to find heterogeneous treatment effects can be problematic. One extreme approach is to continuously run tests to compare arbitrarily selected subgroups, until finding statistically significant evidence for treatment effect heterogeneity over two (or more) subgroups. However, with every statistical test there is always a small probability of obtaining a false positive result (aka Type I error) – that is, a result which incorrectly indicates the presence of a heterogeneous treatment effect when none actually exists. If a large number of tests are performed in an unprincipled manner, there is a high chance of obtaining spurious false positive results(Brookes, Whitley, Peters, Mulheran, Egger and Davey, 2001).

To avoid this issue, a commonly recommended approach is to pre-specify subgroups of interest *before* carrying out the analysis, and then *only* examine those subgroups in the analysis stage. However, this approach is also limited; since we only perform tests on pre-specified subgroups, we may fail to discover variables which turn out to be unexpectedly relevant for modelling heterogeneity, but are not already specified pre-analysis(Brookes, Whitley, Peters, Mulheran, Egger and Davey, 2001).

Both problems can be avoided through machine learning techniques. We use a particular machine learning algorithm called a *causal forest* to estimate heterogeneous treatment effects(Athey, Tibshirani and Wager, 2019). A key advantage of causal forests is that their estimates are consistent and asymptotically Gaussian under certain assumptions[4]; this also allows for the construction of valid confidence intervals(Wager and Athey, 2018).

D. Related literature

Prior studies in the literature broadly find that the ACA Medicaid expansions led to a significant average increase in insurance coverage(Frean, Gruber and Sommers, 2017; Miller and Wherry, 2017). Courtemanche, Marton, Ukert, Yelowitz and Zapata (2017) estimate that the average treatment effect over the whole population was 2.9 percentage points.

In addition, many studies consider treatment effects over particular subgroups, or try to assess heterogeneity in the treatment effect associated with certain covariates. We summarise some important results below.

Several studies(Kaestner, Garrett, Chen, Gangopadhyaya and Fleming, 2017; Wehby and Lyu, 2018) focus their attention on adults with low education, finding that Medicaid expansions led to substantial coverage gains in this subgroup.

Guth, Artiga and Pham (2020) conduct a review of 65 studies, and broadly find evidence that the Medicaid expansions narrowed (but did not eliminate) racial disparities in insurance coverage. This suggests that minorities responded more strongly to the expansions than non-minorities.

Several studies examine heterogeneity in the treatment effect by age, but s6s appear mixed. Courtemanche, Marton, Ukert, Yelowitz and Zapata (2017) and Wehby and Lyu (2018) find that coverage gains were largest for those aged 18-34; however, Frean, Gruber and Sommers (2017) and Dworsky, M. and Eibner, C. (2016) instead find that older adults were more responsive.

There are also some limitations to the prior literature, which we address in this study. Firstly, many prior studies that estimate heterogeneous treatment effects, e.g. Courtemanche, Marton, Ukert, Yelowitz and Zapata (2017), have limited discussion of why such information would be relevant to policymakers. The present study addresses this shortcoming by explicitly linking estimates of treatment effect heterogeneity to their implications for policy targeting.

Secondly, we are not aware of any prior studies using causal forests to estimate heterogeneous treatment effects of the ACA Medicaid expansions on insurance coverage. Causal forests possess several advantages over more traditional approaches, as described in Section II C, and our use of this novel methodology is a key contribution of this study. Indeed, our causal forest discovered that private for-profit employment was a highly relevant covariate for modelling heterogeneity, as will be discussed in Section V F. This variable was not generally mentioned in prior studies; thus, our novel methodology contributes new evidence to the existing literature.

III. METHODOLOGY

Section III A formally defines treatment effects through the potential outcomes framework, and Section III B discusses our estimating equation. Section III C gives a detailed overview of causal forests, and describes how they estimate heterogeneous treatment effects. Finally, Section III D discusses estimation of average treatment effects.

A. Potential outcomes framework

Let us consider a setup with N observations indexed $i = 1, \ldots, N$. Each observation i is associated with covariates X_i, outcome Y_i and treatment indicator W_i. W_i is a dummy variable where $W_i = 1$ if i has received the treatment, and $W_i = 0$ if not.

Under the Neyman-Rubin potential outcomes framework(Rubin, 2005), denote by $Y_i(1)$ the potential outcome for i had they received the treatment, and denote by $Y_i(0)$ the potential outcome for i had they not received the treatment. Notice that this relies on a Stable Unit Treatment Value Assumption (SUTVA), namely that the

[4] See Section III C 4.

potential outcomes for a given unit are only dependent on its own treatment assignment, and not on how the treatment is assigned to other units.

The causal effect of the treatment on each unit then is defined as $Y_i(1) - Y_i(0)$. However, the observed outcome for a unit is $Y_i = Y_i(0) + W_i[Y_i(1) - Y_i(0)]$, that is, we only ever observe one of the potential outcomes. This is the fundamental problem of causal inference – since we do not observe both potential outcomes, we cannot identify individual-level treatment effects(Holland, 1986).

Instead, we consider the following average estimands. First, the Average Treatment Effect (ATE):

$$E[Y_i(1) \; - \; Y_i(0)], \tag{1}$$

By default, we will refer to the ATE over the whole population as simply 'the ATE'. However, we could also compute an average treatment effect over all the individuals belonging to a particular subgroup. We will distinguish this latter concept by referring to it as a 'subgroup ATE' throughout.

A second key estimand is the Conditional Average Treatment Effect (CATE):

$$E[Y_i(1) \; - \; Y_i(0)|X_i = x], \tag{2}$$

which computes the average treatment effect specifically for individuals with covariates x. The CATE is of key interest, since it allows us to estimate heterogeneous treatment effects across individuals with different covariates.

To identify these average estimands, we impose two further assumptions(Weeks and Christiansen, 2020). The first is unconfoundedness, i.e. that potential outcomes $Y_i(1)$ and $Y_i(0)$ are independent of the treatment status W_i after conditioning on X_i:

$$Y_i(0), Y_i(1) \perp W_i|X_i, \tag{3}$$

The second assumption is overlap, which requires that each observation has some probability of being treated and of not being treated. Formally, for all i,

$$0 < Pr(W_i = 1|X_i = x) < 1 \quad \forall x. \tag{4}$$

To illustrate why overlap is required for identification, consider an example from Weeks (2022) where we have a single categorical covariate X_i taking three values $\{1, 2, 3\}$. Suppose that all three values of X_i have treated observations, but only $X_i = 1$ and $X_i = 3$ have untreated observations. This violates overlap for $X_i = 2$. Subsequently, treated observations with $X_i = 2$ do not have good counterfactuals in the control group, meaning we cannot identify the treatment effect when $X_i = 2$.

B. Estimating equation, orthogonalisation and the R-learner

We focus on the CATE, since we are interested in estimating heterogeneous treatment effects[5]. In what follows we will notate the CATE as:

$$\tau(x) = E[Y_i(1) \; - \; Y_i(0)|X_i = x], \tag{5}$$

We then seek to estimate the following equation:

$$Y_i = f(X_i) + \tau(X_i)W_i + \epsilon_i, \tag{6}$$

where $f(X_i)$ is the expected outcome in the absence of treatment. Eq. (6) will ultimately be estimated non-parametrically by causal forests, hence we do not need to impose strong functional form assumptions on $f(X_i)$ and $\tau(X_i)$. This is a key advantage of a machine learning approach, as noted in Section II C.

Before applying the causal forest algorithm, a crucial intermediate step is to transform the estimating Eq. (6) via 'orthogonalisation', following (Robinson, 1988). Define the marginal outcome $m(x) = E(Y_i|X_i = x)$, and notate the propensity score $e(x) = E(W_i|X_i = x) = Pr(W_i = 1|X_i = x)$. We can then rewrite the original estimating Eq. (6) in the following orthogonalised form(Kreif, Diaz-Ordaz, Moreno-Serra, Mirelman, Hidayat and Suhrcke, 2022):

$$Y_i - m(X_i) = \tau(X_i)[W_i - e(X_i)] + \epsilon_i, \tag{7}$$

This is similar in spirit to the Frisch-Waugh-Lovell theorem, in that we have 'partialled out' the effect of X_i on Y_i and W_i. The advantage of Eq. (7) is that the causal forest can purely focus on modelling $\tau(X_i)$, while estimation of $e(x)$ and $m(x)$ can be carried out separately by any off-the-shelf predictive machine learning algorithm(Nie and Wager, 2021)[6]. By contrast, had the causal forest estimated the original Eq. (6), it would have 'wasted' some computational resources on modelling $f(X_i)$ as well as $\tau(X_i)$; this would then reduce accuracy in the final CATE estimates(GRF Labs, 2022).

Nie and Wager (2021) refer to this orthogonalised approach as the 'R-learner'. They note that $\tau(.)$ can be estimated by minimising the expected squared-error loss function for the orthogonalised Eq. (7):

$$\begin{aligned}\hat{\tau}(x) = & argmin_\tau \Bigg\{ E\Bigg(\Big[(Y_i - m(X_i)) - \\ & (W_i - e(X_i))\tau(X_i) \Big]^2 \mid X_i = x \Bigg) \Bigg\} \quad \forall x,\end{aligned} \tag{8}$$

[5] Furthermore, once we have estimated the CATE we can also use this to estimate ATEs. See Section III D.

[6] Our implementation of causal forests estimates $e(x)$ and $m(x)$ separately using an 'honest' version of a random forest. Random forests are defined in Section III C 1 and honesty in Section III C 2.

If we take the first-order condition of Eq. (8) and rearrange it, we can then obtain the associated local moment condition:

$$E\Big((Y_i - m(X_i))(W_i - e(X_i)) - \tau(X_i)(W_i - e(X_i))^2 \mid X_i = x\Big) = 0 \quad \forall x. \tag{9}$$

We term Eq. (9) the 'R-learner moment condition'; note that it can be used to identify the CATE $\tau(.)$. This means the causal forest algorithm can estimate the CATE based on the R-learner moment condition. We will detail this in Sections III C 3 and III C 4.

C. GRF and causal forests

This section is structured into four parts, which build step-by-step towards a description of causal forests. To lay some groundwork, Section III C 1 gives an introduction to regression trees and random forests, and Section III C 2 discusses honest estimation. Section III C 3 then gives an overview of the Generalised Random Forest algorithm (GRF), illustrating similarities and differences with standard random forests. Finally, Section III C 4 discusses how causal forests can be implemented as a special case of GRF.

1. *Regression trees and random forests*

Consider a regression problem where we seek to predict an outcome Y_i given covariates X_i. In other words, we wish to estimate $E(Y_i|X_i)$. We can achieve this through an algorithm called a regression tree.

In general, a tree is a partitioning of the covariate space[7] into subregions called leaves(Athey and Imbens, 2016). Consider a simple example of a tree from Hastie, Tibshirani and Friedman (2017). Here we have two covariates, X_1 and X_2, and the tree partitions the covariate space into 5 leaves as shown in Fig. 1.

The tree is constructed using a 'training sample'; this is a sample of data points which are called 'training examples'. Each leaf R is then associated with a prediction $\bar{Y}_R$, which is the average outcome of the training examples falling into that leaf.

After constructing the tree, we can get a prediction for a new data point which was *not* in the original training sample. We refer to this new data point as a 'test point', and denote its covariates as x. To get a prediction for x, we see which leaf x lands in and return the prediction associated with that leaf.

The tree partitioning is constructed by a process called recursive binary splitting. To begin with, we divide the whole covariate space into two subregions, based on a particular splitting variable and a split point for that variable. For instance, we might take X_1 as the splitting variable, and t_1 as a split point. Then we define two subregions, one containing observations with $X_1 \leq t_1$, and another containing observations with $X_1 > t_1$.

The algorithm chooses the split that minimises a certain target criterion called the splitting criterion. Here, the splitting criterion is the mean squared prediction error of the new tree that we obtain after making the split(Weeks, 2022).

The same binary splitting procedure is then recursively applied to our two subregions to obtain four subregions, and so on. Splitting continues until further splits would cause the resulting subregions to be smaller than some pre-specified minimum node size. The splitting procedure is 'greedy' in that it simply makes the best available split at each step, rather than selecting splits that might lead to better trees in a future step.

The sequence of splits for our example tree is shown in Fig. 1. The root node of a tree is the node where the first split is made – in Fig. 1, this would be the node where we split at $X_1 = t_1$. The depth of a given node is then defined as the number of edges linking that node back to the root node. For instance, the node where we split at $X_2 = t_4$ would have a depth of 2.

An advantage of a single tree is its interpretability, since we can easily see the hierarchical structure of splits that were selected to partition the covariate space. This hierarchical structure also captures dependencies and interactions between variables. For instance, we can see in Fig. 1 that $X_2 = t_2$ is only a relevant split point for data points that also have $X_1 \leq t_1$, i.e. data points that went to the left-hand side of the tree after the first split.

A disadvantage of using single regression trees is that they typically exhibit high variance. To improve stability in the estimates, we can instead use a 'forest-based' method where we draw B random subsamples from the data[8], and grow a different tree on each subsample. Then, to get a prediction for a test point x, we start by pushing x down each tree. For the b^{th} tree, we see which leaf node x lands in and obtain an associated prediction $\hat{\mu}_b$. The final forest prediction simply averages over all the individual tree predictions, so $\hat{\mu} = \sum_{b=1}^{B} \hat{\mu}_b$(Tay, 2020).

One issue with a forest-based approach is that if one of the covariates is a strong predictor, almost all the individual trees will select this covariate as the first splitting variable. This makes the trees highly correlated, in which case averaging them together does not lead to a large stability improvement(Weeks, 2022). Hence,

[7] The covariate space is the set of all possible values that could be taken by the covariates.

[8] In theory these subsamples could be drawn either with or without replacement. We implemented our estimation procedure using the R package grf, which draws the samples *without* replacement(GRF Labs, 2022)

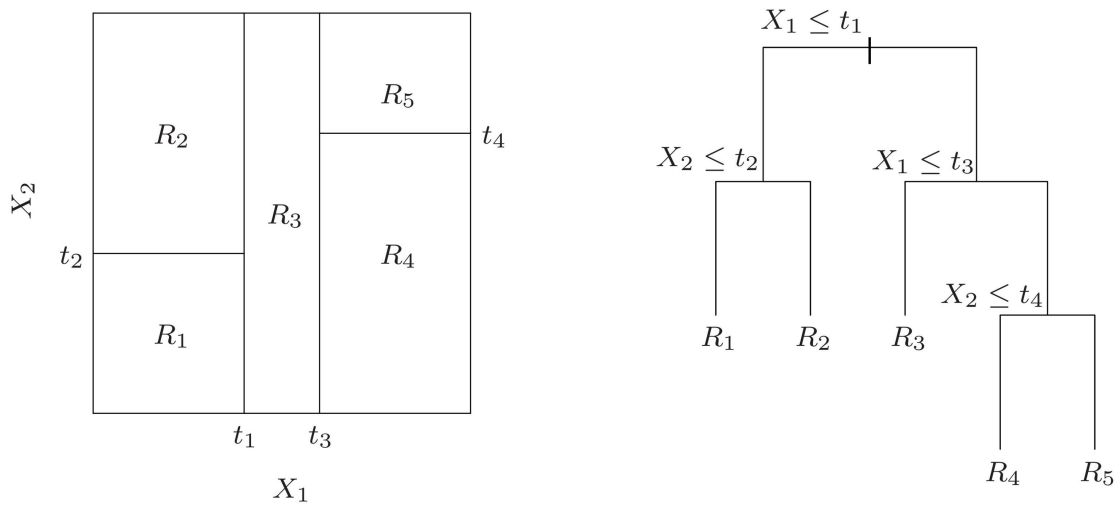

FIG. 1. Left panel: A partitioning of the covariate space. Right panel: An example of a tree. Reproduced from Hastie, Tibshirani and Friedman (2017), p.306, Figure 9.2.

Breiman (2001) proposes a modified forest-based procedure known as *random forests*, where at each splitting point a random subset of variables is made available to split on. This aims to reduce the correlation between trees as much as possible.

2. *Honest estimation*

Standard random forests use the same training data to construct partitions and to generate the estimates associated with each leaf. However, note that the algorithm will typically group data points with spuriously extreme Y_i into the same leaves(Athey and Imbens, 2016). The estimates generated for these leaves will then exhibit bias, due to them being computed over data points which were selected on the basis of having extreme outcomes.

To avoid this source of bias and ensure consistent estimates, we want to use separate data for constructing partitions and generating leaf estimates. We therefore impose a condition called 'honesty': for every training example i, the outcome Y_i is either used to place the splits, or to generate leaf estimates, but not both(Wager and Athey, 2018).

To implement honest estimation in practice, we randomly divide the training sample in half and only use the first half to partition the covariate space. After this is completed, the first half of the training sample is discarded; the second half is then used to populate the leaves and subsequently generate the leaf estimates(GRF Labs, 2022).

3. *Overview of GRF*

We now turn to an overview of the Generalised Random Forest (GRF) algorithm[9], contrasting it with the standard random forest algorithm described in the previous section. Understanding GRF is of key importance, since we will ultimately implement causal forests as a special case of GRF (see Section III C 4).

Consider a setup where we have N training examples indexed $i = 1, \ldots, N$. Furthermore, let each training example i be associated with some observable quantity O_i along with covariates X_i. In the case of a simple prediction problem, as with random forests, O_i simply comprises the outcome variable i.e. $O_i = \{Y_i\}$. If instead we are interested in estimating a treatment effect, O_i would comprise the outcome and treatment assignment variables, i.e. $O_i = \{Y_i, W_i\}$.

Now suppose we are interested in a parameter $\theta(x)$. The GRF algorithm allows us to estimate $\theta(.)$ at any given point $X_i = x$, so long as $\theta(.)$ is identified by a local moment condition[10]. To introduce some notation, we represent the local moment condition as(Athey, Tibshirani and Wager, 2019):

$$E(\psi_{\theta(x)}(O_i)|X_i = x) = 0 \ \ \forall x, \tag{10}$$

To estimate $\theta(x)$ empirically for a given x, we convert Eq. (10) into a weighted sample moment and minimise this with respect to $\theta(x)$. The weights are constructed

[9] This overview draws extensively on Athey, Tibshirani and Wager (2019).

[10] In the context of our study, the parameter of interest is the CATE, and it is identified by the R-learner moment condition Eq. (9). See Section III C 4.

such that training examples with covariates similar to x are assigned greater weight, because they are more informative about the true x.

Let us denote the weight on each training example X_i as $\alpha_i(x)$. The minimisation problem can then be expressed as(Athey, Tibshirani and Wager, 2019):

$$\hat{\theta}(x) \in argmin_{\theta(x)} \left\{ \left\| \sum_{i=1}^{N} \alpha_i(x) \psi_{\theta(x)}(O_i) \right\|_2 \right\}, \quad (11)$$

Solving Eq. (11) yields the GRF estimate for $\theta(x)$. Note that in the special case that this expression has one unique root, we can equivalently estimate $\theta(x)$ by just solving(Athey, Tibshirani and Wager, 2019):

$$\sum_{i=1}^{N} \alpha_i(x) \psi_{\theta(x)}(O_i) = 0, \quad (12)$$

To calculate weights $\alpha_i(x)$, we begin by growing B trees indexed $b = 1, \ldots, B$. In doing so, some core elements of standard random forests are preserved: each tree is grown on a random subsample of the data using recursive binary splitting, and at each splitting point a random subset of variables is made available to split on. However, we do not use the same splitting criterion as a standard random forest. Instead, the GRF splitting criterion is designed such that, intuitively speaking, we will ultimately increase heterogeneity in the estimates of $\theta(x)$ as fast as possible[11].

For each tree b, let $L_b(x)$ denote the set of training examples falling in the same leaf as x. Then the weights $\alpha_i(x)$ represent the frequency with which the i-th training example lands in the same leaf as x, calculated across all the trees. This can be represented as follows(Athey, Tibshirani and Wager, 2019):

$$\alpha_{bi}(x) = \frac{\mathbf{1}(\{X_i \in L_b(x)\})}{|L_b(x)|}, \quad \alpha_i(x) = \frac{1}{B} \sum_{b=1}^{B} \alpha_{bi}(x). \quad (13)$$

This usage of trees to obtain weights, rather than a set of estimates to be averaged together, is substantially different in perspective to other standard forest-based algorithms such as random forests.

There is only one special case where the two perspectives are equivalent: this is in fact the scenario previously discussed in Section III C 1, where we seek to predict an outcome Y_i given covariates X_i. Here, the weighting-based GRF procedure is mathematically equivalent to the averaging-over-predictions approach of a standard random forest(Athey, Tibshirani and Wager, 2019).

However, unlike standard random forests, GRF can be applied to a much broader range of tasks than a simple prediction problem: it can estimate *any* parameter identified by an appropriate local moment condition. In these more general scenarios the weighting- and averaging-based approaches are no longer equivalent, and it is the weighting-approach which proves much more effective, as stressed by Athey, Tibshirani and Wager (2019).

It is instructive to think of GRF as similar to an adaptive nearest neighbour method, insofar as we seek a weighted set of neighbours for test point x(Athey, Tibshirani and Wager, 2019). Importantly, with GRF we allow the neighbourhood size to vary depending on how the trees are constructed, especially since the leaf nodes may potentially have different sizes. This is in contrast to a classical k-nearest neighbours method, where the neighbourhood size k is a pre-determined parameter.

4. *Heterogeneous treatment effect estimation with causal forests*

Causal forests can be implemented as a special case of GRF. To do so, we begin by rewriting the R-learner moment condition Eq. (9) as:

$$E\Big(\underbrace{(Y_i - m(X_i))(W_i - e(X_i)) - \tau(X_i)(W_i - e(X_i))^2}_{\psi_\tau} \mid X_i = x \Big) = 0 \quad \forall x, \quad (14)$$

Recall also that the GRF estimator could be obtained by solving Eq. (12):

$$\sum_{i=1}^{N} \alpha_i(x) \psi_{\theta(x)}(O_i) = 0,$$

We take ψ_τ from Eq. (14) and plug this into Eq. (12), by setting $\psi_{\theta(x)}(O_i) = \psi_\tau$. After rearranging, the GRF estimate for the CATE can be written as(Athey and Wager, 2019):

$$\hat{\tau}(x) = \frac{\sum_{i=1}^{N} \alpha_i(x)(Y_i - \hat{m}(X_i))(W_i - \hat{e}(X_i))}{\sum_{i=1}^{N} \alpha_i(x)(W_i - \hat{e}(X_i))^2}. \quad (15)$$

As discussed in Section III A, the fundamental problem of causal inference implies that individual-level treatment effects cannot be identified. This poses difficulties in evaluating the performance of our CATE estimator Eq. (15) since, unlike in predictive problems, the 'ground truth' of the causal effect is not observed(Athey and Imbens, 2016). In response, the literature has developed a series of theoretical results showing that causal forest estimates are consistent and asymptotically Gaussian under a certain set of assumptions. An important assumption is that trees are honest in the sense defined in Section III C 2(Athey, Tibshirani and Wager, 2019). We also require unconfoundedness Eq. (3) and overlap Eq. (4) to hold(Wager and Athey, 2018).

[11] For a technical description of the GRF splitting criterion see Sections II B and II C of Athey, Tibshirani and Wager (2019).

D. Average treatment effect estimation

The CATE is highly granular, since it is a function of every single covariate. To get a higher-level perspective on the treatment effects, it is often helpful to aggregate the estimated CATEs to obtain estimates of the average treatment effect (ATE) over selected subgroups or the whole population.

Naively, we can accomplish this by simply averaging the CATE estimates over all the observations in our population or subgroup of interest. However, a more accurate estimate can be obtained by using a 'doubly-robust' estimator(GRF Labs, 2022). This estimator is constructed using estimates for the propensity score and conditional mean outcomes. So long as *either* of these two estimates are consistent, the doubly-robust estimator will then deliver a consistent estimate of the ATE. Another key advantage of the doubly-robust estimator is that it attains the semiparametric efficiency bound(Farrell, 2015).

In this context, the doubly-robust estimator takes the following form(Athey and Wager, 2019; GRF Labs, 2022)[12]:

$$\hat{\tau} = \frac{1}{N}\sum_i \left\{ \hat{\tau}(X_i) + \frac{W_i - \hat{e}(X_i)}{\hat{e}(X_i)(1-\hat{e}(X_i))}(Y_i - \hat{\mu}(X_i, W_i)) \right\}, \tag{16}$$

where $\hat{\mu}(X_i, W_i)$ is defined as:

$$\hat{\mu}(X_i, W_i) = \hat{m}(X_i) + \hat{\tau}(X_i)[W_i - \hat{e}(X_i)], \tag{17}$$

Note that $\hat{\mu}(X_i, W_i)$ is an estimate of the conditional mean outcome $E(Y_i|X_i, W_i)$. This can be seen by rearranging the orthogonalised estimating Eq. (7) to obtain:

$$Y_i = \underbrace{m(X_i) + \tau(X_i)[W_i - e(X_i)]}_{\mu(X_i, W_i)} + \epsilon_i.$$

Note also that the doubly robust estimator is asymptotically Gaussian, allowing us to construct valid confidence intervals(Wager, 2020).

IV. IDENTIFICATION STRATEGY AND DATA

A. Identification strategy

In the context of our study, we can rewrite the estimating Eq. (6) as follows:

$$Y_{is} = f(X_{is}) + \tau(X_{is})W_{is} + \epsilon_{is}. \tag{18}$$

where the indexes indicate individual i in state s. Y_{is}, the outcome variable, is a dummy which equals 1 if i has health insurance and 0 if not. W_{is} indicates the treatment status, which equals 1 if state s has expanded Medicaid and 0 if not. X_{is} contains a set of individual- and state-level covariates.

To ensure unconfoundedness, it is necessary to control for any variables that might influence the treatment assignment W_{is} and also separately influence the outcome variable Y_{is}. Our identification strategy assumes that unconfoundedness holds after controlling for the following three state-level observables:

1. *State political affiliation.* The debate around the ACA was highly partisan, with Democrats typically supporting the ACA while Republicans opposed it. Many Republican-led states therefore opted out of the Medicaid expansion as part of their broader opposition to the ACA as a whole.

   State political affiliation may have also had a separate influence on individual insurance statuses. Motivated by a conservative ideology of individual responsibility, some Republican states have modified their Medicaid programs to feature increased 'cost-sharing'(Baker and Hunt, 2016). This reduces the proportion of healthcare costs covered by insurance, thereby requiring individuals to fund more healthcare expenses by themselves. Since insurance programmes are less generous in Republican states, fewer individuals in these states may be incentivised to enrol in them.

2. *State expenditure per capita.* States had to partially fund Medicaid expansions, and several opted out because they were unwilling to do so(Sommers and Epstein, 2011). For instance, states may have been averse to large state expenditures for political or ideological reasons, or else they may have lacked financial resources. Any of these factors should also be associated with lower state expenditure per capita.

   In addition, lower state expenditure per capita may have separately influenced individual insurance statuses. For instance, this may have meant fewer state-sponsored advertisement campaigns to inform individuals about insurance options, making individuals less likely to obtain insurance. Indeed, Karaca-Mandic, Wilcock, Baum, Barry, Fowler, Niederdeppe and Gollust (2010) find that state-sponsored advertisements significantly increased Medicaid enrolment in 2014.

3. *State vs federal insurance exchanges.* The ACA established private insurance exchanges where individuals and small businesses could directly purchase health insurance. States were given the option to run these exchanges themselves or else rely on a federally-run exchange. Whether exchanges were state- or federally-run may correlate with the decision to opt out of Medicaid expansions, e.g. states lacking interest or administrative capacity

[12] We have omitted additional 'cross-fitting' notation for clarity of exposition.

might have opted out of Medicaid expansions and also relied on the federal exchange(Courtemanche, Marton, Ukert, Yelowitz and Zapata, 2017).

Furthermore, the choice of state- or federally-run exchanges separately influenced individual insurance enrolment(Terrizzi, Mathews-Schultz, and Deegan, 2022). State-run exchanges may have operated at reduced administrative burden, since they could focus on the population of a single state, rather than needing to function at a federal level. This arguably increased the efficiency of state-run exchanges in stimulating insurance uptake.

Note that apart from the Medicaid expansions and insurance exchanges, other major ACA provisions were implemented at a federal level in 2014. As they took effect uniformly in all states, they did not systematically correlate with whether states expanded Medicaid or not, minimising confounding concerns.

B. Data sources

Individual-level data was obtained from the 2014 American Community Survey (ACS). This is a yearly, nationwide survey of over 3 million Americans. Within each state a random sample of individuals is drawn; these samples are then put together to get the overall dataset(US Census Bureau, 2015). Individuals selected to take part in the survey are required to complete it by law, thus minimising any non-response bias(Wehby and Lyu, 2018).

The ACS records whether each individual has health insurance or not, letting us construct the outcome variable Y_{is}. The ACS also records a rich set of socioeconomic, demographic and health-related covariates for each individual; these were included in X_{is}. The full list of individual-level covariates used in this study is provided in Table I.

Information on which states expanded Medicaid in 2014 was taken from Kaiser Family Foundation (2022). All states that implemented Medicaid expansions in 2014 did so at the start of the year, except for Michigan (which expanded Medicaid in April 2014) and New Hampshire (which expanded Medicaid in August 2014)(Kaiser Family Foundation, 2022). These two states could not be cleanly classified as treated or untreated, since they effectively switched status partway through the year. We therefore dropped them from the study. Reassuringly, key results from other studies of the Medicaid expansions are not significantly affected by keeping or dropping these states(Courtemanche, Marton, Ukert, Yelowitz and Zapata, 2017; Kaestner, Garrett, Chen, Gangopadhyaya and Fleming, 2017).

Table II indicates our state control variables and their data sources. These variables could not be obtained for the District of Columbia (D.C.), which is a federal district and not a state, so we dropped D.C. from our study.

One of our state controls was obtained by grouping states into sextiles, based on their level of state expenditure per capita (see Table II for details). We did this to ensure overlap, as each sextile contained a mixture of treated and untreated states. By contrast, had we directly used state expenditure per capita as a control variable, this would have violated overlap. For example, state expenditure per capita was \$6,790 for New York. But all individuals in New York experienced the Medicaid expansion, hence all observations with a state expenditure per capita of \$6,790 would have received the treatment and none of these observations would have been untreated.

Along similar lines, we obtained another state control by grouping states into terciles, based on the percentage of Republican individuals in each state. Again, this was done to preserve overlap, since we could ensure that each tercile contained a mixture of treated and untreated states.

In theory our study could have used cross-sectional data from a later year than 2014, thereby lengthening the time window between the initial Medicaid expansions and our observed dataset. This would have allowed us to capture longer-term treatment effects of the Medicaid expansions. However, we believed that this would not provide much extra information, since we expected insurance coverage to respond quite quickly to Medicaid expansions. Insurance enrolment is itself a reasonably fast process; moreover, the Medicaid expansions were announced several years prior to 2014, meaning that information about the new policies should have already diffused through the population in advance.

Furthermore, a risk is that over longer time windows, it is much harder to measure and control for confounders. For instance, categorising states by political affiliation is very difficult over longer time periods: some states like Pennsylvania and Louisiana had Republican state governors in 2014, but switched to Democrat governors in subsequent years(Ballotpedia, 2022), making it ambiguous how to classify them overall.

These considerations motivated our choice of 2014 cross-sectional data for this study.

C. Adjustments for clustering

In our study, the treatment is assigned at the state-level. As Abadie, Athey, Imbens and Wooldridge (2022) argue, this means we should adjust our estimation procedure to account for state-level clustering. We therefore made two key modifications to our causal forest algorithm, following Athey and Wager (2019), which we discuss below.

In the standard causal forest algorithm, each tree is grown on a random subsample of the training data. However, Athey and Wager (2019) propose a cluster-robust modification to this sampling procedure which proceeds in two stages: in the first stage a subsample of clusters is randomly drawn, and then in the second stage individ-

TABLE I. Individual-level covariates that were used in this study, from the 2014 American Community Survey(US Census Bureau, 2007).

Continuous variables	
Age	Supplementary security income past 12 months
Interest, dividends, net rental income past 12 months	Wage or salary income past 12 months
Public assistance income past 12 months	Total income
Retirement income past 12 months	All other income past 12 months (e.g. unemployment or disability compensation)
Self-employment income past 12 months	
Dummy variables. All variables below equal 1 if the individual has the characteristic of interest, and 0 if not	
Male	State government employee
Self-care difficulty	Federal government employee
Hearing difficulty	Self-employed in own non-incorporated business, practice or farm
Vision difficulty	Self-employed in own incorporated business, practice or farm
Independent living difficulty	Working without pay in family business or farm
Ambulatory difficulty	Civilian employed, at work
Veteran-service connected disability	Civilian employed, with a job but not at work (e.g. due to parental leave or illness)
Cognitive difficulty	Unemployed
Any type of disability	Not in labour force
Married	Works in management
Widowed	Works in business-related industry
Divorced	Works in finance
Separated	Works in computer- or maths-related industry
Never married or under 15	Works in engineering
Gave birth to child in past 12 months	Works in science
Grandparents living with grandchildren	Community worker
Female with children aged under 6 years only	Legal worker
Female with children aged 6 to 17 years only	Educational worker
Female with children, some aged under 6 and others aged 6 to 17	Entertainment worker
Female with no children	Medical worker
US citizen by being born in the US	Works in healthcare support
US citizen by being born in Puerto Rico, Guam, US Virgin Islands, or Northern Marianas	Works in protective services (e.g. police or firefighters)
US citizen by being born abroad of American parent(s)	Works in food preparation and serving
US citizen by naturalization	Works in cleaning
Not a citizen of the US	Works in personal care and services (e.g. hairdressers, travel guides)
Speaks a language other than English at home	Works in sales
Lived in same location one year ago	Office administrator or assistant
Did not live in US or Puerto Rico one year ago	Works in farming, fishing or forestry
Lived in Puerto Rico one year ago and now lives in US	Works in construction
Recently migrated from one part of the US to another	Works in extraction (e.g. mining)
American Indian or Alaska Native	Works in repair, installation and maintenance
Asian	Works in production (e.g. factory worker)
Black	Transportation worker
Native Hawaiian	On active duty in military
Other Pacific Islander	Used to be on active duty in military, but not now
Some other race	Only on active duty for training in Reserves/National Guard
White	Never served in military
Employee of private for-profit company or individual	Has degree in science or engineering
Employee of private not-for-profit, tax-exempt, or charitable organization	
Local government employee	

Categorical variables	
Highest level of education attainment. Takes the following values:	Quarter of birth. Takes the following values:
1: No schooling completed	1: January - March
2: Nursery school or preschool	2: April - June
3: Kindergarten	3: July - September
4: Grade 1	4: October -December
5: Grade 2	
6: Grade 3	
7: Grade 4	
8: Grade 5	
9: Grade 6	
10: Grade 7	
11: Grade 8	
12: Grade 9	
13: Grade 10	
14: Grade 11	
15: Grade 12 , but no high school diploma	
16: High school diploma	
17: GED or alternative credential	
18: Some college, but less than 1 year	
19: At least one year of college credit but no degree	
20: Associate's degree	
21: Bachelor's degree	
22: Master's degree	
23: Professional degree beyond Bachelor's degree	
24: Doctorate degree	

TABLE II. Description and data sources for state controls[a].

State control	Description	Data source
Whether insurance exchanges were state- or federally-run	Dummy variable. Equals 1 if insurance exchange was state-run and 0 if federally-run.	Giovannelli and Lucia (2015)
Sextiles by state expenditure per capita	Categorical variable, taking values 1-6. Equals 1 if state ranks in the bottom sextile by state expenditure per capita, and so on, until equalling 6 if state ranks in top sextile by state expenditure per capita.	Underlying data on state expenditure per capita taken from Ballotpedia (2022)
Terciles by percentage of Republicans individuals in state	Categorical variable, taking values 1-3. A measure of state political affiliation. Equals 1 if state ranks in the bottom tercile by percentage of Republicans, and so on, until equalling 3 if state ranks in top tercile by percentage of Republicans.	Underlying data on percentage of Republicans in state taken from the Gallup poll in Jones (2015). Note that we classified individuals as Republican if they reported either strong identification or a leaning towards the Republican Party.
Whether state governor was Republican	Dummy variable. A measure of state political affiliation. Equals 1 if state governor was Republican and 0 if not.	Ballotpedia (2022)
Whether upper house of state legislature was majority Republican	Dummy variable. A measure of state political affiliation. Equals 1 if upper house of state legislature was majority Republican and 0 if not.	Ballotpedia (2022)
Whether lower house of state legislature was majority Republican	Dummy variable. A measure of state political affiliation. Equals 1 if lower house of state legislature was majority Republican and 0 if not.	Ballotpedia (2022)

[a] *Note: for the last two measures, a complication came from Nebraska, whose state legislature contains a single house. This house was majority Republican in 2014. Since Nebraska's single house combines the functions of what would normally be an upper and lower house, we simply recorded Nebraska as having a majority-Republican upper and lower house.*

ual data points are randomly sampled from each selected cluster. This modification reduces overfitting, and ensures that standard errors and confidence intervals are cluster-robust.

Athey and Wager (2019) also propose a cluster-robust modification to the doubly-robust ATE estimator discussed in Section III D. Under this modification, we first compute a doubly-robust estimate of the ATE within each cluster. Then an overall estimate of the ATE is obtained by a simple averaging of all the within-cluster ATE estimates.

V. RESULTS

We trained a causal forest with 500 trees and a minimum node size of 5000, using the R package grf.

Firstly, we confirmed that the Medicaid expansions had a positive overall effect on insurance coverage, with an estimated ATE over the whole population of 2.41 percentage points. This estimate is comparable to earlier studies, such as Courtemanche, Marton, Ukert, Yelowitz and Zapata (2017).

In what follows, we delve into heterogeneity in our estimated treatment effects. Section V A visualises and tests the overall extent of heterogeneity. Sections V B to V F explore which variables were most relevant for modelling heterogeneity, allowing us to characterise which subgroups were most responsive to Medicaid expansions.

A. Testing for heterogeneity in the treatment effects

Recall that we motivated our study by arguing that evidence on the CATEs would facilitate targeting of future Medicaid policy towards particularly responsive individuals. However, this argument is only compelling if there *is* indeed heterogeneity in the treatment effects. If instead all individuals responded identically, this would no longer provide grounds to target one individual over another.

Therefore, we are interested in visualising and testing the extent of heterogeneity in our estimated CATEs. To begin, we rank all individuals in the population by the magnitude of their estimated CATEs, from largest to smallest. In other words, the individual with the largest estimated CATE receives the highest ranking, and so on. Now consider a subgroup containing the individuals who are ranked in the top tenth of the population by this measure.

If there is strong heterogeneity in the treatment effect, we should expect the ATE over this subgroup to be substantially larger than the ATE over the whole population. We define the difference between these two ATEs as the Targeting Operator Characteristic (TOC)(GRF

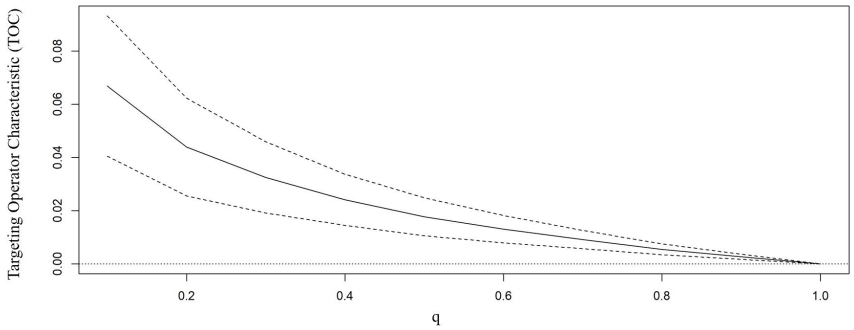

FIG. 2. The estimated TOC curve for our study. Dotted lines indicate 95% confidence intervals.

Labs, 2022).

More generally, we could compute the TOC for the subgroup of individuals ranked in the top q of the population, for any fraction q. This lets us plot a 'TOC curve', with q on the x axis, and the associated TOC for each q on the y axis. Fig. 2 shows the estimated TOC curve for our study[13].

Note that at $q = 1$, we are subtracting the ATE among individuals in the the top 100% of the population from the ATE over the whole population. Evidently both terms are equivalent, so the TOC at $q = 1$ will equal zero.

Furthermore, consider a case where there was *no* heterogeneity in the treatment effect, i.e. the CATE was the same for all individuals. This would mean that the ATE among individuals in the top q of the population would equal the ATE of the whole population, for any value of q. Therefore, the TOC would equal 0 everywhere. In Fig. 2 we can see visually that the TOC curve is quite far from zero over most of its range, suggesting our estimated CATEs exhibit strong heterogeneity.

We can formally test for heterogeneity in our estimated CATEs as follows. First, we estimate the average TOC over all values of q. We then test if this estimate is significantly above zero[14]; if so, this provides strong evidence in favour of heterogeneity in our estimated treatment effects. In our study this estimate was indeed significantly above zero at all conventional levels[15], strongly supporting the presence of heterogeneity in our estimated CATEs.

[13] This was implemented using the R package grf. Following default settings, the curve is plotted for $q = (0.1, 0.2, \ldots, 1.0)$.

[14] The estimate is a sample average, so we can derive a central limit theorem for it and construct hypothesis tests. See Yadlowsky, Fleming, Shah, Brunskill and Wager (2021).

[15] The point estimate was 0.02727, with a standard deviation of 0.005138, to 4 significant figures.

B. Variable importance

We focus next on assessing which variables were most relevant for modelling heterogeneity; this is captured by measures of 'variable importance'. Such information will subsequently allow us to characterise the subgroups that were most responsive to the Medicaid expansions.

A simple measure of variable importance for some given covariate x_j is as follows[16]: take a weighted sum over the frequency with which x_j is split on at each depth of the causal forest, up to a maximum depth of 4. The exact formula for this measure is(O'Neill and Weeks, 2019):

$$importance(x_j) = \frac{\sum_{k=1}^{4} \left[\frac{\sum_{all\ trees} N_{x_j}}{\sum_{all\ trees} N_{total}} \right] k^{-2}}{\sum_{k=1}^{4} k^{-2}}. \tag{19}$$

where k^{-2} is the weight for depth level k, N_{x_j} is number of splits on x_j at depth k, N_{total} is total number of splits at depth k. A high value of $importance(x_j)$ would indicate that x_j was very relevant for modelling heterogeneity, since a high proportion of trees chose to split on it.

One drawback of measure Eq. (19) is that it does not capture dependencies and interactions between variables. Such information can be seen easily from the hierarchical structure of single trees, as discussed in Section III C 1. However, this information is lost when using many trees together in causal forests. This is an important limitation of variable importance measures and of forest-based methods more broadly.

Another issue with measure Eq. (19) is that it exhibits some bias towards continuous variables or categorical variables with high cardinality. This is because such variables have more potential splitting points, reflecting the fact that they embody more information(O'Neill and

[16] This is the default variable importance measure in the R package grf.

TABLE III. Top 5 covariates, by variable importance.

Ranking	All covariates		Individual level covariates only	
	Variable importance	Variable	Variable importance	Variable
1	0.1610	Age	0.1610	Age
2	0.1181	Sextiles by state expenditure per capita	0.1129	Total income
3	0.1129	Total income	0.0698	Whether in private for-profit employment
4	0.0698	Whether in private for-profit employment	0.0543	Highest level of educational attainment
5	0.0543	Highest level of educational attainment	0.0530	Whether speaking a language other than English at home

Weeks, 2019). Altmann, Toloşi, Sander and Lengauer (2010) instead propose a bias-corrected 'permutation importance' measure – however, this involves fitting causal forests hundreds or thousands of times, which we lacked the computational resources to implement. Reassuringly, O'Neill and Weeks (2019) find that both measures give a similar ranking of variable importance.

The five most important covariates from our causal forest are reported in Table III. One of our state controls, sextiles by state expenditure per capita, was among these top five; more broadly, all our state controls had fairly high importance. This is reassuring since it means the algorithm made strong use of the state controls, improving robustness to confounding. However, we are otherwise most interested in which *individual*-level covariates were strongly associated with treatment effect heterogeneity; we ultimately want to target Medicaid policies to different individuals, not different states. To this end, we also enumerate the five most important *individual*-level covariates; again, see Table III.

Notice that total income had high variable importance. This is fairly self-explanatory since, by design, the Medicaid expansions were targeted to those below a certain income threshold (138% of the federal poverty line). Regardless, this result is still useful as a sanity check – had the causal forest failed to pick up on income, we might be worried that it was not properly modelling treatment effect heterogeneity.

In the following sections, we consider the four other individual covariates with high variable importance: age, educational attainment, speaking a non-English language at home, and private for-profit employment. Overall, our findings suggest that particularly responsive individuals were typically aged 25-34; had a high school diploma as their highest level of educational attainment; spoke a language other than English at home, and/or were in private for-profit employment.

C. Age

To examine age-related heterogeneity in the treatment effect, we divided the population into seven subgroups, defined by the following age brackets: 0-17, 18-24, 25-34, 35-44, 45-54, 55-64, and over 65. Estimated subgroup ATEs are plotted in Fig. 3 and reported in Table IV.

We find that the age bracket 25-34 had the largest subgroup ATE. Conway (2020) notes that this age bracket also exhibits the highest health uninsurance rates. He argues that many individuals in this age range have recently dropped out of their parents' insurance coverage, since young adults in the US can only remain dependent on their parents' health plans until the age of 26. However, since most 25-34 year-olds tend to be in good health, they may not be willing to pay much money to enrol in a new insurance plan. Hence many of these individuals would rather remain uninsured, unless given access to affordable public health insurance.

The uninsured rate steadily falls for older age brackets(Conway, 2020); these individuals typically start to incur greater health problems, so they may have already obtained insurance outside of Medicaid. Treatment effects of Medicaid expansions therefore peak in the 25-34 age bracket.

D. Educational attainment

To examine education-related heterogeneity, we divided the population into five subgroups. The highest level of educational attainment in each subgroup was respectively:

1. Less than primary school
2. More than primary school but less than having graduated from high school
3. A high school diploma or equivalent
4. At least some college education, up to having completed an undergraduate degree
5. A postgraduate degree

Estimated subgroup ATEs are plotted in Fig. 3 and reported in Table IV. The subgroup ATE was highest for category 3, i.e. for those whose highest educational attainment was a high school diploma or equivalent.

Most individuals with less education than a high school diploma are children. However, many children in low-income families were already eligible for Medicaid pre-expansion, so they were not as strongly impacted by the

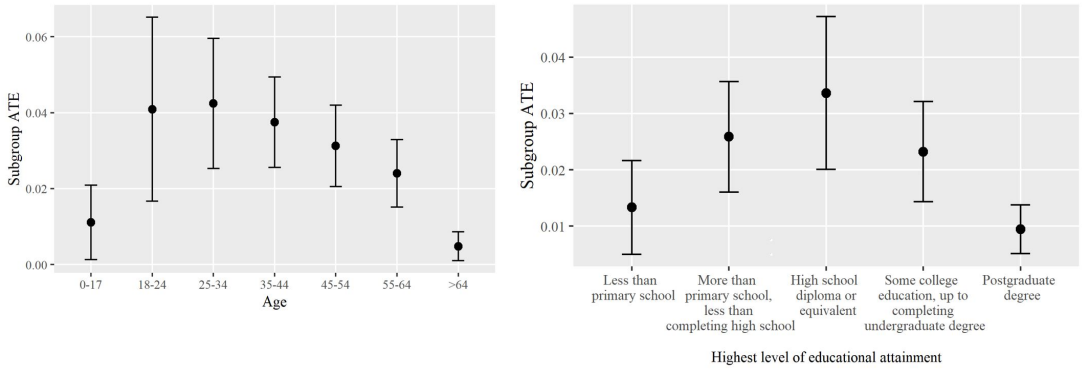

FIG. 3. Estimated subgroup ATEs by age (left panel) and educational attainment (right panel). Plotted with 95% confidence intervals.

TABLE IV. Estimated subgroup ATEs.

Subgroup	Estimated subgroup ATE (to 4 decimal places)	Standard error (to 5 decimal places)
Aged 0-17	0.0111	0.00501
Aged 18-24	0.0409	0.01236
Aged 25-34	0.0425	0.00875
Aged 35-44	0.0375	0.00609
Aged 45-54	0.0313	0.00547
Aged 55-64	0.0240	0.00453
Aged 65 or older	0.0048	0.00193
Less than primary school education	0.0133	0.00425
More than primary school education but less than having graduated high school	0.0259	0.00502
Highest level of education is a high school diploma or equivalent	0.0336	0.00691
At least some college education, up to completing an undergraduate degree	0.0232	0.00454
Has a postgraduate degree	0.0095	0.00219
Speaks a language other than English at home	0.0336	0.00985
Does not speak a language other than English at home	0.0220	0.00432
In private for-profit employment	0.0346	0.00698
Not in private for-profit employment	0.0177	0.00333

expansions themselves. By contrast, many individuals at a high school diploma level are working adults, of whom a significant proportion became Medicaid-eligible for the first time due to the expansions. This may explain why treatment effects are higher on average for individuals at a high school diploma level, compared to individuals who have not yet graduated high school.

On the other end of the educational spectrum, university graduates typically earn higher incomes than high school graduates, as well as being more likely to work in jobs that already offer extensive health insurance benefits(US Institute of Medicine, 2001). This might explain why university graduates are less responsive to Medicaid expansions, as compared to high school graduates.

E. Speaking a language other than English at home

We split the population into two subgroups: those that spoke a language other than English at home, and those that spoke English at home. Estimating the subgroup ATEs revealed that on average, the former subgroup responded more strongly to the Medicaid expansions; see Table IV.

Those that speak a language other than English at home usually belong to a racial, cultural or linguistic minority within the US. Individuals in minority groups typically have lower incomes, making it hard for them to afford private insurance. Furthermore, they are less likely to work in sectors offering health insurance benefits(Artiga, Hill, Orgera and Damico, 2021). Both factors mean that minorities rely more extensively on affordable public health insurance, making them more responsive to the Medicaid expansions.

Moreover, many minorities have reduced cultural, lin-

guistic or institutional familiarity with the US, meaning they face particularly strong administrative and informational barriers to insurance enrolment(Stuber, Maloy, Rosenbaum and Jones, 2000). Such barriers were, however, reduced by Medicaid expansions. For instance, many states that expanded Medicaid also sponsored advertisement campaigns to improve information about health insurance options(Karaca-Mandic, Wilcock, Baum, Barry, Fowler, Niederdeppe and Gollust, 2010). This may have helped stimulate a large response to Medicaid expansions among minorities.

F. Private for-profit employment

The estimated subgroup ATE for those in private for-profit employment was 3.46 percentage points. This was higher than the ATE over the whole population, which was 2.41 percentage points. The higher-than-average responsiveness of those in private for-profit employment could stem from the fact that employer-provided coverage in the private sector has been steadily decreasing in amount and rising in cost(Claxton, Larry and Damico, 2016). As private-sector employees struggle to obtain affordable coverage from their employers, many are increasingly enrolling in public insurance programmes instead(Strane, Kanter, Matone, Glaser and Rubin, 2019), hence their strong reaction to an expansion in Medicaid eligibility.

This was not a variable which we had anticipated to be important pre-analysis. Indeed, earlier studies of the Medicaid expansions[17] did not consider this variable, so our discovery of it is an important contribution to the literature. This demonstrates the value of a machine learning approach: we were able to remain agnostic pre-analysis about which variables were important, while letting the algorithm *discover* such variables for us during the analysis phase.

VI. CONCLUSION

In this paper we estimated the heterogeneous treatment effects of Medicaid expansions on insurance coverage, using an observational study based on 2014 cross-sectional data. We emphasised that estimating *heterogeneous* treatment effects let us determine which individuals were most responsive to the expansions. Such individuals could be targeted by future Medicaid policies, so this evidence is of key interest for policymakers.

We found strong overall evidence of heterogeneity in our estimated CATEs. We also found that particularly responsive individuals were typically aged 25-34; had a high school diploma as their highest level of educational attainment; spoke a language other than English at home, and/or were in private for-profit employment. Hence, one potential targeted Medicaid policy could be an expansion of eligibility to individuals with incomes below 138% of the federal poverty line, *provided* that they belong to one of the categories above.

Our usage of causal forests to estimate CATEs is a key contribution of this paper. The advantage of this approach over more traditional methods is exemplified by our discovery that private for-profit employment was a highly important variable. This variable was missed by prior studies that relied on pre-specified subgroup analyses, demonstrating how novel machine learning approaches can bring new evidence to the existing literature.

There are some limitations to our study. Firstly, our identification strategy assumed that unconfoundedness held after controlling for state political affiliation, state expenditure per capita, and whether insurance exchanges were state- or federally-run. We have argued in favour of this approach given the institutional context of the ACA, but it is never fully possible to rule out the possibility of unobserved confounders which we failed to control for.

Another point to acknowledge is that, while our estimates of heterogeneous treatment effects would be a crucial piece of evidence in designing targeted Medicaid programmes, they would not be the only factor to consider. In particular, there are also complex ethical issues in designing targeted programmes, which we have mostly left outside the scope of this paper. Notably, our results suggest targeting Medicaid towards minority subgroups, particularly those not speaking English at home. This would essentially form an 'affirmative action' policy, which remains a contentious subject of ethical debate in the US. Proponents of affirmative action argue that it redresses the historic oppression of minorities; meanwhile, critics view it as an unjust form of reverse discrimination(Chemerinsky, 1996).

An interesting extension to this paper would be to take a stance on the social welfare function, accounting for potential equity concerns such as the debate on affirmative action. We might also wish to consider heterogeneous costs of programme implementation(Davis and Heller, 2020) – for instance, it might be more expensive to target minority subgroups who face informational barriers to Medicaid enrolment, as doing so effectively would require states to spend more on advertising campaigns. By combining these elements with our existing evidence on heterogeneous treatment effects, we could carry out a complete social cost-benefit analysis to select the optimal targeted policy. One paper taking this approach is Knittel and Stolper (2021); in their particular case study, they find that policy targeting based on causal forest estimates delivers significantly higher net social benefits than other targeting methods. This is an exciting avenue for future research, and it could yield rich insights for policymakers seeking to design socially optimal targeted policies.

Cha, A. and Cohen, R. (2022): "Demographic Variation in Health Insurance Coverage: United States, 2020," *National Health Statistics Reports*, 169, 1-15.
Papanicolas, I., Woskie, L. R. and Jha, A. K. (2018): "Health Care Spending in the United States and Other High-Income Countries," *Journal of the American Medical Association*, 319(10), 1024-1039.
Coleman, D. L.(2006): "Viewpoint:The Impact of the Lack of Health Insurance: How Should Academic Medical Centers and Medical Schools Respond?," *Academic Medicine*, 81, 728-731.
Garthwaite, C., Graves, J.,d Gross, T., Karaca, Z., Marone, V. and Notowidigdo, M.(2019): "All Medicaid Expansions Are Not Created Equal: The Geography and Targeting of the Affordable Care Act," *Brookings Papers on Economic Activity*, 1-70.
Center for Medicaid and CHIP Services(2022): "Medicaid," Available at: https://www.medicaid.gov/medicaid/index.html.
Gruber, J. and Sommers, B. D.(2019): "The Affordable Care Act's Effects on Patients, Providers, and the Economy: What We've Learned so Far," *Journal of Policy Analysis and Management*, 38(4), 1028-1052.
Courtemanche, C., Marton, J., Ukert, B., Yelowitz, A. and Zapata, D.(2017): "Early Impacts of the Affordable Care Act on Health Insurance Coverage in Medicaid Expansion and Non-Expansion States," *Journal of Policy Analysis and Management*, 36(1), 178-210.
Hellmann, J.(2019): "GOP Georgia governor proposes limited Medicaid expansion," Available at: https://thehill.com/policy/healthcare/468865-republican-georgia-governor-proposes-limited-medicaid-expansion/.
Alker, J. and Brooks, T.(2022): "Millions of Children May Lose Medicaid: What Can Be Done to Help Prevent Them From Becoming Uninsured?," Report at Georgetown University Health Policy Institute.
Corlette, S. and Kona, M.(2022): "Mitigating Coverage Loss When the Public Health Emergency Ends," Available at: https://www.commonwealthfund.org/blog/2022/mitigating-coverage-loss-when-public-health-emergency-ends.
Brookes, S. T., Whitley, E., Peters, T. J., Mulheran, P.A., Egger, M. and Davey-Smith, G.(2001): "Subgroup analyses in randomised controlled trials: quantifying the risks of false-positives and false-negatives," *Health Technology Assessment*, 5, 1-56.
Buettgens, M. and Green, A.(2022): "What Will Happen to Medicaid Enrollees' Health Coverage after the Public Health Emergency?: Updated Projections of Medicaid Coverage and Costs," Report at Urban Institute.
Athey, S., Tibshirani, J. and Wager, S.(2019): "Generalized random forests," *The Annals of Statistics*, 2, 1148-1178.
Wager, S. and Athey, S.(2018): "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests," *Journal of the American Statistical Association*, 113(523), 1228-1242.
Frean, M., Gruber, J. and Sommers, B. D.(2017): "Premium subsidies, the mandate, and Medicaid expansion: Coverage effects of the Affordable Care Act," *Journal of Health Economics*, 53, 72-86.
Miller, S. and Wherry, L. R.(2017): "Health and Access to Care during the First 2 Years of the ACA Medicaid Expansions," *New England Journal of Medicine*, 376(10), 947-956.
Kaestner, R., Garrett, B., Chen, J., Gangopadhyaya, A. and Fleming, C.(2017): "Effects of ACA Medicaid Expansions on Health Insurance Coverage and Labor Supply," *Journal of Policy Analysis and Management*, 36(3), 608-42.
Wehby, G. L. and Lyu, W.(2018): "The Impact of the ACA Medicaid Expansions on Health Insurance Coverage through 2015 and Coverage Disparities by Age, Race/Ethnicity, and Gender," *Health Services Research*, 53(2), 1248-1271.
Guth, M., Artiga, S. and Pham, O(2020): "Effects of the ACA Medicaid Expansion on Racial Disparities in Health and Health Care," Report at Kaiser Family Foundation.
Dworsky and Eibner (2016): "The Effect of the 2014 Medicaid Expansion on Insurance Coverage for Newly Eligible Childless Adults," Santa Monica, CA, US: RAND Corporation.
Rubin, D. B. (2005): "Causal Inference Using Potential Outcomes," *Journal of the American Statistical Association*, 100(469), 322-331.
Holland, P. W.(1986): "Statistics and Causal Inference," *Journal of the American Statistical Association*, 81(396), 945-960.
Weeks, M. and Christiansen, T. G.(2020): "Understanding the Distributional Aspects of Microcredit Expansions," Available at: http://arxiv.org/abs/2011.10509.
Weeks, M.(2022): "Topic 6: Treatment Effects and Policy Evaluation. Part 2B, Paper 10.," *Lecture Notes, University of Cambridge*.
Robinson, P. M.(1988): "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56(4), 931-954.
Kreif, N., Diaz-Ordaz, K, Moreno-Serra, R., Mirelman, A., Hidayat, T. and Suhrcke, M.(2022): "Estimating heterogeneous policy impacts using causal machine learning: a case study of health insurance reform in Indonesia," *Health Services and Outcomes Research Methodology*, 22(2), 192-227.
Nie, X. and Wager, S.(2021): "Quasi-oracle estimation of heterogeneous treatment effects," *Biometrika*, 108(2), 299-319.
GRF Labs(2022): "The GRF Algorithm, " Available at: https://grf-labs.github.io/grf/REFERENCE.html.
Sommers, B. D. and Epstein, A. M.(2011): "Why States Are So Miffed about Medicaid — Economics, Politics, and the "Woodwork Effect"," *New England Journal of Medicine*, 365(2), 100-102.
Daoud, A. and Johansson, F.(2019): "Estimating Treatment Heterogeneity of International Monetary Fund Programs on Child Poverty with Generalized Random Forest," SocArXiv.
Athey, S. and Imbens, G.(2016): "Recursive partitioning for heterogeneous causal effects," *Proceedings of the National Academy of Sciences of the United States of America*, 113(27), 7353-7360.
Hastie, T. and Tibshirani, R. and Friedman, J.(2017): "The Elements of Statistical Learning: Data Mining, Inference, and Prediction," Berlin, Germany: Springer.
Tay, K.(2020): "What are general-

[17] See Section II D.

ized random forests?," Available at: https://statisticaloddsandends.wordpress.com/2020/06/05/what-are-generalized-random-forests/.

Weeks, M.(2022): "Topic 9b and 9c: Regression and Causal Forests. R301b.," *Lecture Notes, University of Cambridge*.

Breiman, L.(2001): "Random Forests," *Machine Learning*, 45(1), 5-32.

Athey, S. and Wager, S.(2019): "Estimating Treatment Effects with Causal Forests: An Application," *Observational Studies*, 5(2), 37-51.

Farrell, M.(2015): "Robust Inference on Average Treatment Effects with Possibly More Covariates than Observations," *SSRN Scholarly Paper*.

GRF Labs(2022): "Estimating conditional means," Available at: https://grf-labs.github.io/grf/articles/muhats.html.

Wager, S.(2020): "STATS 361: Causal Inference," *Lecture Notes, Stanford University*.

Baker, A. and Hunt, L. (2016): "Counterproductive Consequences of a Conservative Ideology: Medicaid Expansion and Personal Responsibility Requirements," *American Journal of Public Health*, 106(7), 1181-1187.

Karaca-Mandic, P., Wilcock, A., Baum, L., Barry, C. L., Fowler, E. F., Niederdeppe, J. and Gollust, S. E. (2010): "The Volume Of TV Advertisements During The ACA's First Enrollment Period Was Associated With Increased Insurance Coverage," *Health Affairs*, 36(4), 747-754.

Terrizzi, S., Mathews-Schultz, A. L. and Deegan, M. M.(2022): "State versus federal health insurance marketplaces: A bigger deal for Medicaid and a smaller deal for the individual mandate," *Health Policy OPEN*, 3, 100059.

US Census Bureau(2007): "PUMS Accuracy of the Data (2014)," Available at: https://www2.census.gov/programs-surveys/acs/tech_docs/pums/accuracy/2014AccuracyPUMS.pdf.

Kaiser Family Foundation(2022): "Status of State Medicaid Expansion Decisions: Interactive Map," Available at: https://www.kff.org/medicaid/issue-brief/status-of-state-medicaid-expansion-decisions-interactive-map/.

Giovannelli, J. and Lucia, J.(2015): "The Experiences of State-Run Marketplaces That Use HealthCare.gov," Available at: https://www.commonwealthfund.org/publications/issue-briefs/2015/sep/experiences-state-run-marketplaces-use-healthcaregov.

Ballotpedia(2022): "State government trifectas," Available at: https://ballotpedia.org/State_government_trifectas.

Ballotpedia(2022): "Total state government expenditures," Available at: https://ballotpedia.org/Total_state_government_expenditures.

Jones, J.(2015): "Massachusetts, Maryland Most Democratic States, " Available at: https://news.gallup.com/poll/181475/massachusetts-maryland-democratic-states.aspx.

Abadie, A., Athey, S., Imbens, G. and Wooldridge, J.(2022): "When Should You Adjust Standard Errors for Clustering?," arXiv.

US Census Bureau(2015): "2014 ACS PUMS Data Dictionary," Available at: https://www2.census.gov/programs-surveys/acs/tech_docs/pums/data_dict/PUMSDataDict14.pdf.

GRF Labs(2022): "Assessing heterogeneity with RATE," Available at: https://grf-labs.github.io/grf/articles/rate.html.

Yadlowsky, S., Fleming, S., Shah, N., Brunskill, E. and Wager, S.(2021): "Evaluating Treatment Prioritization Rules via Rank-Weighted Average Treatment Effects," arXiv.

O'Neill, E. and Weeks, M.(2019): "Causal Tree Estimation of Heterogeneous Household Response to Time-Of-Use Electricity Pricing Schemes," arXiv.

Altmann, A., Toloşi, L., Sander, O. and Lengauer, T.(2010): "Permutation importance: a corrected feature importance measure," *Bioinformatics*, 26(10), 1340-1347.

Conway, D.(2020): "Adults Age 26 Had Highest Uninsured Rate Among All Ages, Followed By 27-Year-Olds," Available at: https://www.census.gov/library/stories/2020/10/uninsured-rates-highest-for-young-adults-aged-19-to-34.html.

US Institute of Medicine(2001): "Who Goes Without Health Insurance? Who Is Most Likely to Be Uninsured?," *Coverage Matters: Insurance and Health Care*, Washington D.C.,US: National Academies Press (US).

Artiga, S., Hill, L., Orgera, K. and Damico, A.(2021): "Health Coverage by Race and Ethnicity, 2010-2019," Available at: https://www.kff.org/racial-equity-and-health-policy/issue-brief/health-coverage-by-race-and-ethnicity/.

Stuber, J., Maloy, K., Rosenbaum, S. and Jones, K.(2000): "Beyond Stigma: What Barriers Actually Affect the Decisions of Low-Income Families to Enroll in Medicaid?," *Report at Health Policy and Management Issue Briefs*.

Claxton, G.,d Larry, L. and Damico, A.(2016): "Diminishing Offer and Coverage Rates Among Private Sector Employees," Available at: https://www.kff.org/private-insurance/issue-brief/diminishing-offer-and-coverage-rates-among-private-sector-employees/.

Strane, D., Kanter, G. P., Matone, M., Glaser, A. and Rubin, D. M.(2019): "Growth Of Public Coverage Among Working Families In The Private Sector," *Health Affairs*, 38(7), 1132-1139.

Chemerinsky, E.(1996): "Making Sense of the Affirmative Action Debate," *Ohio Northern University Law Review*, 22, 1159-1176.

Davis, J. M.V. and Heller, S. B.(2020): "Rethinking the Benefits of Youth Employment Programs: The Heterogeneous Effects of Summer Jobs," *The Review of Economics and Statistics*, 102(4), 664-677.

Knittel, C.R. and Stolper, S.(2021): "Using Machine Learning to Target Treatment: The Case of Household Energy Use," ASSA 2021 Virtual Annual Meeting, American Economic Association.