Heterogeneous Treatment Effects and Efficient Policy Learning: Evidence from the Oregon Health Experiment[☆]

Shishir Shakya[1]

Abstract

One of the primary objectives of any field experiments and quasi-experiments is to understand whom to treat. But, identical policy intervention distinctly affects individuals and subpopulations. Policymakers prefer to learn heterogeneous treatment effects to understand the underlying mechanisms that drive the results and to design targeted policy. I use the Oregon health insurance experiment data to estimate heterogeneous treatment effects of insurance on health outcomes. Then, I propose insurance assignment strategies constraining for ethical, legislative, and political reasons that can maximize the overall welfare.

Keywords:

JEL codes:

[☆]While retaining full responsibility for errors and omissions, I thank Jane Ruseski, Jousha C. Hall, Feng Yao, Sultan Altruki, Edwardo Minucci and participants of YSI North America Convening DoubleTree by Hilton and USC Dornsife, Los Angeles held in Feb 22–24, 2019.

[1]Shishir Shakya, College of Business and Economics & Regional Research Institute (RRI), West Virginia University, Morgantown WV 26506 E-mail: shishir.shakya@mail.wvu.edu

1. Introduction

This study has two objectives. First, to estimate the heterogeneous treatment effects of insurance on health care utilization (including primary and preventive care as well as hospitalizations), out-of-pocket medical expenditures and medical debt, and self-reported physical and mental health. Second, to propose insurance assignment strategies constraining for ethical, legislative, and political reasons that maximizes the health care utilization (including primary and preventive care as well as hospitalizations), lower out-of-pocket medical expenditures and medical debt (including fewer bills sent to collection), and better self-reported physical and mental health. I utilize the Oregon Health Insurance Experiment dataset and several causal inference theory grounded machine learning approaches.

An extensive literature studying the impact of insurance coverage on the health outcomes report average treatment effects. But, claiming causality is tenacious due to endogeneity. Endogeneity arises because it is difficult to control for observed and unobserved confounding variables among the insured and uninsured population (Levy & Meltzer, 2008). For example, a comparison of the health between those with and without health insurance, (say the Medicaid) can reveal that Medicaid is bad for one's health (Baicker & Finkelstein, 2011). Because people with poor health are more likely to get insurance compare to healthy people.

Random assignment of insurance can avoid such confounding problems (Finkelstein et al., 2012) and Oregon's health insurance lottery provides a rare opportunity to estimate the causal effects of being allowed to apply for Medicaid (Baicker & Finkelstein, 2011). However, unlike numerous papers(see Allen, Baicker, Finkelstein, Taubman, & Wright, 2010; Baicker et al., 2013; Baicker, Allen, Wright, & Finkelstein, 2017; Baicker, Finkelstein, Song, & Taubman, 2014; Baicker & Finkelstein, 2011; Finkelstein et al., 2012; Grossman et al., 2016; Taubman, Allen, Wright, & Baicker, 2014; Zhou, Baicker, Taubman, & Finkelstein, 2017) which provide average treatment effects estimates, this paper ventures to estimate the heterogeneous treatment effect of having insurance and uses lottery selection as an instrument for insurance coverage.

Exploring treatment heterogeneity can help to quantify the size of effects on different subpopulations, which is valuable to improve program targeting and to understand the underlying mechanisms that drive the results. Therefore, I intend to estimate treatment heterogeneity of insurance on several health outcomes. To investigate effect heterogeneity, one can stratify the data

in mutually exclusive groups or include interactions in a regression (Athey & Imbens, 2017). But performing ad hoc searches –implementing data mining or p-hacking to discover for particularly responsive subgroups– may lead to false discoveries or may mistake noise for a true treatment effect (Davis & Heller, 2017). While with a preregistered analysis plan, a researcher can commit in advance to study only a subgroup. Such a preregistered analysis plan can hedge against researchers to claim the data mining or p-hacking results but may also prevent from discovering unanticipated results and developing new hypotheses.

This paper implements the "causal trees" approach of Athey & Imbens (2015) and "causal forest" of Wager & Athey (2017) which explore treatment heterogeneity in a principled way. These methods conceptually align with the potential outcome framework or Rubin causal model (Holland, 1986; Imbens & Rubin, 2015a; Rubin, 1974) but modify the classification and regression tree (CART)– a machine learning algorithm of Breiman et al. (1984)– to implement propensity score matching in high-dimensional functions of covariates rather than just in particular or ad-hoc subgroups. Athey & Imbens (2015) modifications of the CART–a supervised machine learning algorithm, helps to select sub-populations to identify heterogeneity treatment effects; estimate more precise average treatment effects and; test the hypotheses about the difference between the impact in different subpopulations whereas Wager & Athey (2017) uses random forest algorithms of Breiman (2001) to construct asymptotic confidence intervals for the true treatment effect. In short, Athey & Imbens (2015) causal tree and Wager & Athey (2017) causal forest methods are causal inference theory grounded machine learning approaches.

The heterogeneous treatment effect estimations explore how individuals and each subpopulation can be differently susceptible to the identical policy intervention. After learning treatment heterogeneity, more interesting avenue would be to discuss the learning treatment assignment policy or mappings from individual characteristics to treatment assignments. Learning the treatment assignment strategies or learning whom to intervene is essential to policymakers and is a ubiquitous problem in applied economics. For example, whom to serve in youth employment programs (Davis & Heller, 2017), whom to allocate Medicare funding for hip or knee replacement surgery (Kleinberg et al., 2015), who are allocate job training, job search and other assistance based on the National Job Training Partnership Act (Kitagawa & Tetenov, 2018).

I use the efficient policy learning strategies of Athey & Wager (2018) to learn targeted insurance assignment policy rule or strategies that are optimal and feasible after exogenously constraining for ethical, legislative, and political reasons. The treatment assignment problem rarely arises in an unconstrained environment (Athey & Wager, 2018) because if feasible treating everyone in the population is welfare maximizing (Kitagawa & Tetenov, 2018). However, treatment is often expensive. For example –given the fact that through the Affordable Care Act (ACA), the federal government would pay the full cost of expansion through 2016 and would always pay at least 90 percent of the cost of covering the newly-insured population (Norris, 2018)– in January 2014, Oregon expanded and adopted Medicaid expansion known as Oregon Health Plan (OHP) to everyone with incomes up to 133 percent of poverty via the Affordable Care Act (ACA). However, as a part of the deal was always that the federal would gradually reduce their payments and as a result, the state budget of Oregon (~$74 billion for 2017-2019) has about $1 billion budget hole mainly due to the cost of health care (Foden-Vencil, 2018). As a temporary fix, the Oregon legislature proposed a bill called Oregon Measure 101, the Healthcare Insurance Premiums Tax for Medicaid Referendum, which was on the ballot in Oregon as a veto referendum on January 23, 2018 (Ballotpedia, 2018) It was approved. It raises the money via taxing 0.7% on the hospital revenue and 1.5% on the insurers and providers (Foden-Vencil, 2018). However, this bill will repeal in August 2020.

Because the treatments are expensive, learning treatment assignment policy is, therefore, crucial to the policymaker. But, ad-hoc assignment of treatment or assigning treatments to individuals in the target population can be unfair, unethical, illegal, and unpolitical. Therefore, exogenously constraints on covariates like race and gender, etc., I implement efficient policy learning strategies of Athey & Wager (2018) to learn treatment assignment policy rule that depends only on particular types of covariates. Their method uses a doubly robust estimator (Chernozhukov, Escanciano, et al., 2018) for inverse propensity weighting to derive a policy rule. Following Kitagawa & Tetenov (2018), these treatment assignment policy rule are designed from an intention-to-treat perspective because the policy maker's problem is a choice of the eligibility criteria and not as a choice of the take-up rate (decided by individuals); hence, this paper is not interested in the treatment effect on compliers.

2. Literature Reviews

How does health insurance affect health? The answer seems obvious, but Levy & Meltzer (2008) review the literature and draws three conclusions. First, the problem of endogeneity makes causal claims tenuous. Second, the papers that establish causal evidence are, particularly within small subgroup populations. For example, the public health insurance reduces mortality among infants and children (Currie & Gruber, 1996b, 1996a; Hanratty, 1996) while for the elderly, public health insurance improves different outcome but not mortality (Card & Maestas, 2008; Finkelstein & McKnight, 2008; McWilliams et al., 2007b, 2007a). Third, the nature of studies is not representative of the broader population, which prohibits generalizing for policy purposes.

Allen et al. (2010) point out three practical design for insurance and health outcomes research: observational studies, quasi-experimental studies, and randomized experiments. The first is the observational studies which comprise the most substantial part of the literature. Observational studies typically use multivariate regression methods to control the observables confounding variables but do not address the issues of unobservable confounders. Observational studies potentially suffer from the endogeneity as "unobservable differences between the insured and the uninsured may drive the observed differences in health outcomes" (Levy & Meltzer, 2004, 2008).

The second set of study exploits natural experiments to evaluate the effect of health insurance on health outcomes. These studies implement techniques like differences-in-differences estimation, regression discontinuity designs, and instrumental variables to relate the exogenous events that result in variation in health insurance coverage– changes that are plausibly unrelated to health and other underlying determinants of health insurance coverage (Levy & Meltzer, 2008). Because health insurance coverage varies in a way that is unrelated to the unobservable factor, the above methods can disentangle the causal effect of insurance on health. However, the results of natural experiments are valid for only specific population groups and therefore, cannot be generalized. For example, several studies show that public health insurance reduces mortality among infants and children (Currie & Gruber, 1996b, 1996a; Hanratty, 1996) while for the elderly, public health insurance does not reduce mortality (Card & Maestas, 2008; Finkelstein & McKnight, 2008; McWilliams et al., 2007b, 2007a). These “one sizes fit all” policy approach is unlikely to be effective because the channels or mechanisms through which having insurance

affects health outcomes may be different for infants and children than they are for elderly adults. These studies identify the effect of health insurance on health outcomes but do not develop evidence of channels or mechanisms.

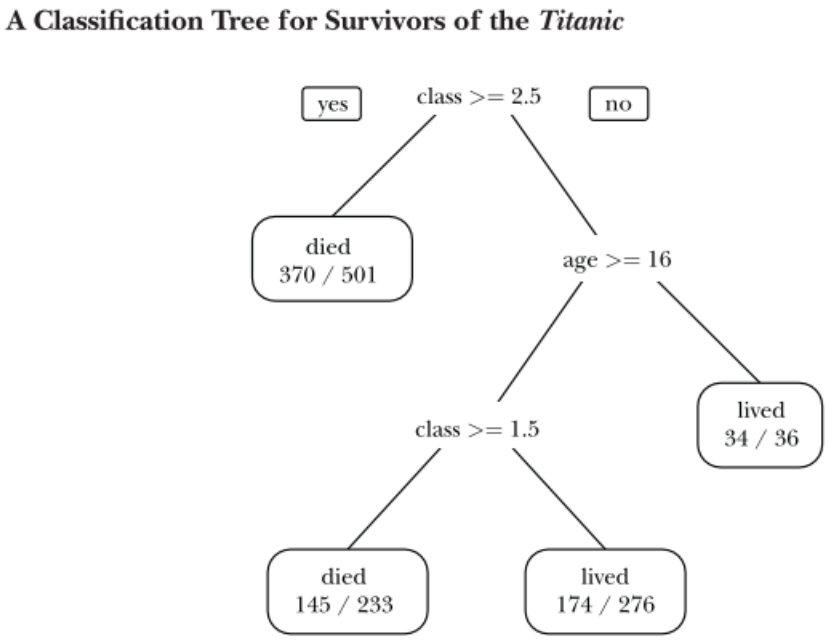
The third set of studies are social experiments, which are the gold standard for establishing causality. The RAND health insurance experiments and the Oregon health insurance experiment are only two of such kind. Newhouse (1994) provides details on the RAND experiment while Finkelstein et al., (2012) and provides details on Oregon experiments. Using RAND experiment data, Newhouse (1994) and Brook et al., (1983) find no significant effect of insurance on the health status of an average adult. Levy & Meltzer (2008) suggest that this finding may be a clue to the fact that the RAND experiment did not randomize people to receive any health insurance; instead, random individuals were treated health insurance with varying degrees of generosity. Finkelstein et al. (2012) find that treatment group has "substantively and statistically significantly higher health care utilization (including primary and preventive care as well as hospitalizations), lower out-of-pocket medical expenditures and medical debt (including fewer bills sent to collection), and better self-reported physical and mental health than the control group".

While observational studies, quasi-experimental studies, and randomized experiments have been dominant in health economics and economics in general. Another strand of literature that modifies the core of machine learning to study causality is gaining momentum recently. Machine learning algorithms behave exceptionally well to make an out-of-sample prediction, and these machine learning algorithms are easy to implement (Mullainathan & Spiess, 2017). However, the machine learning algorithms are not designed to make causal claims (Athey, 2018) and various authors propose a slight modification of "off-the-shelf" machine learning algorithm for causal inference.

I implement the “causal tree,” a machine learning algorithm, as suggested by Athey & Imbens (2015). Athey & Imbens (2015) introduce a data-driven method: to select sub-population to identify heterogeneous treatment effects; to estimate more precise average treatment effects and to test the hypotheses about the difference between the impacts in different subpopulations. Their method makes a comparison between observed outcomes to the counterfactual outcomes under a different regime or treatment which conceptually aligns with the Rubin causal model (Holland, 1986; Imbens & Rubin, 2015a; Rubin, 1974) of potential outcomes framework.

In most economic studies, a large-scale randomize test to perform "A/B tests" may not be feasible and assuredly not in this study. Besides, historically, most datasets have been too small to meaningfully investigate the heterogeneity of treatment effects beyond splitting the sample into several subgroups (Wager & Athey, 2017). The nearest-neighbor matching, kernel methods, and series estimation are classical approaches for nonparametric estimation of heterogeneous treatment effects (see Crump, Hotz, Imbens, & Mitnik, 2008; Lee, 2009; Willke, Zheng, Subedi, Althin, & Mullins, 2012) and performs well with a modest set of covariates. However, these classic approaches collapse when covariates are high dimensional (Wager & Athey, 2017). In an extension of that, studies examining heterogeneity can be purely deceptive when researchers iteratively hunt down the subsets with extreme effects (Assmann et al., 2000; Cook et al., 2004).

Figure-1: An example of a classification and regression tree (CART)



Source: Varian (2014)

Before considering the Athey & Imbens (2015) causal tree method, I begin with its roots – the classification and regression tree (CART) of (Breiman et al., 1984). The CART recursively filters and partitions the large dataset into binary sub-groups (nodes) such that the samples within each subset become more homogenous that fit the response variable. The CART is predictive algorithm is not a causal. However, unlike other machine learning algorithms, the CART generates

decision trees which are easy to transcribe. For example, figure-1 shows features of the Titanic survivors using the CART method. This figure is from Varian (2014).

Furthermore, unlike other statistical methods, the CARTs are free from parametric and structural assumptions (Kim & Koehler, 1995) such as linearity, normality, and equal variance (Horner et al., 2010; Oh & Kim, 2010; Shakya & Lama, 2014) and require little a priori knowledge or theories regarding which variables relationships. The CART can handle nonlinear problems, and are ideal for determining patterns, segmentation, stratifications, predictions and data reductions/screening (Atkins et al., 2007); and deriving models from large, noisy datasets obtained from surveys (Chen et al., 2009). CARTs are less susceptible to outliers and extremes values as well (Shakya & Lama, 2014).

CART handles most of the econometric issues which are partly related to causal effects in experimental or observational studies. The Athey & Imbens (2015) causal tree method modifies the CART algorithm for causal inference. Firstly, using CART, they partition the data into subpopulations such that each partition differs in the magnitude of the treatment effects. Secondly, they estimate treatment effects across each subset of the population. With simulations, they show that their approach enables the construction of confidence intervals for treatment effects, "even in samples with many covariates relative to the sample size, and without "sparsity" assumptions" (Athey & Imbens, 2015). They coined their approach as an "honest" estimation. In their "honest" estimation, one sample is used to construct the partition and another to estimate treatment effects for each subpopulation. Their modified CART algorithm optimizes for the goodness of fit in treatment effects. Unlike the original CART model selection criterion, the causal tree model selection criterion therefore focuses on "improving the prediction of treatment effects conditional on covariates, anticipating that bias will be eliminated by honest estimation, but also accounting for change in the variance of treatment effect estimates within each subpopulation as a result of the split" (Athey & Imbens, 2015).

3. Oregon Health Experiment and data management

3.1 Oregon Health Experiment

In 2008 the Oregon state conducted an extensive public awareness campaign about the lottery opportunity focusing a group of uninsured low-income adults. The program targeted the

adults of ages 19–64 below the federal poverty level (FPL) have assets below $2,000 and not otherwise eligible for public insurance. These adults must be residing in Oregon with legal immigration status or U.S. citizen deprived of health insurance for at least six months. Any eligible person can sign up by telephone, fax, in person sign-up, mail, or online by providing very little demographic information. This program kept the low barrier to sign-up, and no attempts were made to verify the information or screen for program eligibility at sign-up for the lottery.

Among the total of 89,824 individuals who signed up, 35,169 individuals–representing 29,664 households–were selected by lottery. The selected individual in a house is required to submit the appropriate paperwork within 45 days. If they met the eligibility requirement, they enroll in the Oregon Health Plan (OHP), but they must verify their status every six months.

About 60% who were selected by lottery, send back the application and half of those applications failed to meet the requirement therefore about 30% of total selected individuals successfully enrolled in OHP. The OHP Standard provides relatively comprehensive benefits with no consumer cost sharing. It covers physician services, prescription drugs, all significant hospital benefits, mental health, and chemical dependency services (including outpatient services), hospice care, and some durable medical equipment. The OHP excludes the vision and nonemergency dental services. The final study comprises 74922 individuals, and the lottery select 29834 individuals.

3.2 Data management

I use lottery assignment as a treatment indicator $W \in (0,1)$ and the self-reported health outcome in the 12th-month survey as a variable of interest, which ranges from 0 to 5. A higher value represents higher self-reported health quality. For the covariates, I use several demographic variables and health-related variables recorded at the initial survey wave. The purpose of taking health outcome from the 12th-month survey while covariates of initial survey wave are to estimate how treated perceived their health compared to those who are in control group after one year of health insurance experience.

In the Oregon Health Insurance Experiment, three waves of the survey were conducted: baseline, on the 6th month and 12th month. Since the study was not mandatory to fill out; therefore, one primary problem of this dataset is that there exists lots of not available (NA) data entry. For

practical purposes of running the model, I delete observation those with all NAs for all covariates. Then, I impute remaining missing data with Multivariate Imputations by Chained Equations (MICE) algorithm using the random forest. This algorithm is a predictive algorithm to assign NA values with the best approximation. Finally, this study comprises of 26,423 individual's data.

4. Identification Strategy

Consider $i \in \{1,\ldots,N\}$ observations with potential outcomes for each unit to be either $\left(Y_i(0), Y_i(1)\right)$. Following Rosenbaum & Rubin (1983) model, the unit level causal effect is the difference in potential outcomes $\tau_i = Y_i(1) - Y_i(0)$, where, $W_i \in \{0,1\}$ is a binary indicator for the treatment with $W_i = 0$ indicating that unit i received the control treatment and $W_i = 1$ indicating that unit i received the treatment. The realized outcome for unit i is the potential outcomes corresponding to the treatment.

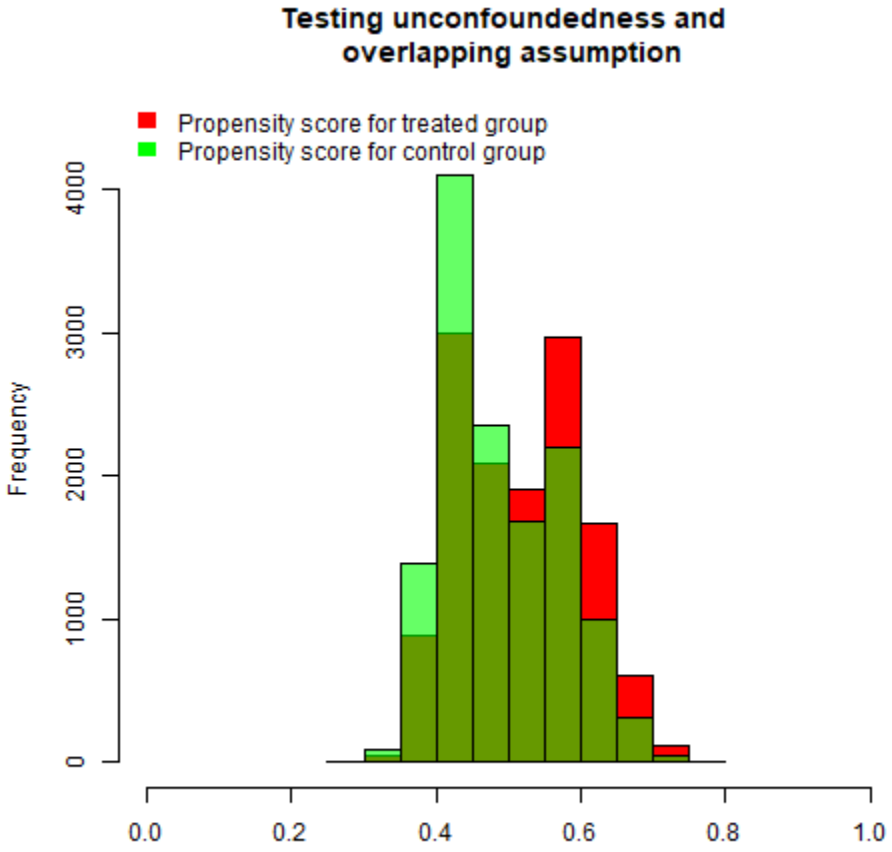
$$Y_i^{obs} = Y_i(W_i) = \begin{cases} Y_i(0) \; if \; W_i = 0, \\ Y_i(1) \; if \; W_i = 1. \end{cases}$$

Let's consider X_i be a K-component vector of features or covariates not to be affected by the treatment. The data consist of triple $\left(Y_i^{obs}, W_i, X_i\right), \forall\, i = 1,\ldots,N$ which are *i.i.d* samples drawn from a large population.

Following Athey & Imbens (2015), this paper assumes: **a)** expectations and probabilities are referring to the distribution induced by the random sampling, or by the (conditional) random assignment of the treatment; **b)** the observations are exchangeable, and that there is no interference (or the stable unit treatment value assumption, or SUTVA (Rubin, 1978) assumption holds), this assumption violets in settings where some units are connected through networks and; **c)** consider $p = pr(W_i = 1)$ be the marginal treatment probability and $e(x) = pr(W_i = 1 \mid X_i = x)$ be the conditional treatment probabilities or the propensity scores as defined by Rosenbaum & Rubin (1983). In a randomized experiment with constant treatment assignment probabilities $e(x) = p$ for all values of x. For causal inference, the strong assumption of randomization conditional on the covariates or "unconfoundedness" (Rubin, 1978) must maintained and formalized as $W_i \perp \left(Y_i(0), Y_i(1)\right) \mid X_i$.

For identification of causal claim, I test for the unconfoundedness and overlap assumption. The unconfoundedness assumption implies treatment is randomly assigned and knowing observable characteristics of an individual i, and their treatment status gives no additional information on the potential outcomes. This means the treatment assignment is independent of the outcome variable $Y_i(1), Y_i(0) \perp W_i \mid X_i$.

Figure-1



If this is the case in the data, the probability of being treated and untreated should be centered around 50 percentages. If this is the case, then the treatment assignment is random and decided for everyone with a coin flip. Given this condition, the overlap assumption guarantees that no subpopulation is entirely located in either the control or treatment groups $\forall x \in \text{supp}(X), 0 < P(W = 1 \mid X = x) < 1.$

Figure 1 shows the distribution of the propensity score. The propensity score is estimated employing L_1 penalized lasso logistic regression technique. The treatment and control propensity scores overlap and center around 0.5 which indicates compliance with the overlapping and unconfoundedness assumption.

In Annex, I present several tables of the mean and the standard deviation of variables among treated and control groups. The table also includes the difference between means and p-values under the null hypothesis if the means between treated and control is zero or not. First, the tables are presented without weighing the observation and then with the inverse probability weighting (See Annex). I find that the mean of variables between control and treated group with inverse probability weighting is statistically the same representing that treated and control individuals are same before the intervention.

5. Athey & Imbens (2015) causal tree

The numerous causal inference studies focus on estimating the population marginal average treatment effect $E\left[Y_i(1)-Y_i(0)\right]$ (Imbens & Rubin, 2015b; Pearl, 2011). However, Athey & Imbens (2015) causal tree estimates the conditional average treatment effect (CATE) $\tau(x) \equiv E\left[Y_i(1)-Y_i(0) \mid X_i = x\right]$. Athey & Imbens (2015) causal tree obtains the estimates of inferences for CATE $\hat{\tau}(\cdot)$ based on the using CART or partitioning the feature spaces that do not vary within the partitions.

A tree starts by considering the entire sample as a single group. For every unique value of each covariate, $X_j = x$, the algorithm forms a candidate data split of the group into two leaves by placing all observation with $X_j \le x$ in a left leaf and all observation with $X_j \ge x$ in a right leaf (Davis & Heller, 2017). It implements split to minimizes some criterion, for example, an in-sample goodness-of-fit test such as mean squared error (MSE) given as $\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2$, where $\hat{y}_i$ is the mean of Y within an observation's leaf. The algorithm repeats the process for each of the two leaves and so on until it reaches a stopping rule. The last leaves are called terminal leaves.

Let's consider a tree $\prod$ to a partition of feature space X, with $\#\left(\prod\right)$ the numbers of elements in the partition given as:

$$\prod = \left\{ \ell_1, \ldots, \ell_{\#(\tau)1} \right\}, \text{ with } \bigcup_{j=1}^{\#(\prod)} \ell_j = \mathrm{X}$$

Let P denote the space of partition. Let $\ell(x;\prod)$ denote the leaf $\ell \in \prod$ such that $x \in \ell$. Let S be the data samples from a population. Let $\pi : \mathrm{S} \mapsto \mathrm{P}$ be an algorithm to construct the partition S and $S \in \mathrm{S}$. Then given a partition $\prod$, for each observation $\left(Y_i^{obs}, X_i, W_i\right)$, the population average condition mean function $\mu(x;\prod)$ is:

$$\mu(w, x;\prod) \equiv \mathrm{E}\left[Y_i(w) \mid X_i \in \ell(x;\prod) \right]$$

And its average causal effect is:

$$\tau(x;\prod) \equiv \mathrm{E}\left[Y_i(1) - Y_i(0) \mid X_i \in \ell(x;\prod) \right]$$

This can be viewed as a step-function approximation to $\mu(x)$. Given sample S the estimated outcome is:

$$\hat{\mu}(w, x; S, \prod) \equiv \frac{1}{\#\left(i \in S_w : X_i \in \ell(x;\prod) \right)} \sum_{i \in S_w : X_i \in \ell(x;\prod)} Y_i$$

which is an unbiased estimate of $\mu(x;\prod)$. This estimator is index by the sample. In the simplest term, $\hat{\mu}(x, S, \prod)$ is conditional mean which is the average outcomes on each of the leaves of the partition data. The estimated causal effect is the difference of treated mean and control mean in the leaf ℓ where it belongs,

$$\hat{\tau}(x; S, \prod) \equiv \tau(\ell) = \hat{\mu}(1, x; S, \prod) - \hat{\mu}(0, x; S, \prod)$$

Before we move forward, let's define some more notations. The notation S represents a data sample drawn from the population. Then, S^{tr}, S^{te} and S^{est} denote a training sample, testing sample and estimation sample. The S_{treat} and $S_{control}$ denote the subsample of treated and control unit. Then, N^{tr}, N^{te} and N^{est} indicate the number of observations in training, testing and estimation sample. N_{treat} and $N_{control}$ denote the number of observations in the treated and control units.

Given a partition $\prod$, Athey & Imbens (2015) modify the standard mean square error (MSE) criteria as $MSE_\tau\left(S^{te}, S^{est}, \prod\right)$ or the mean squared error for treatment effect, where the average treatment effect over a test sample S^{te} and the conditional average treatment is estimated on an estimation sample S^{est} and given as:

$$MSE_\tau\left(S^{te}, S^{est}, \prod\right) \equiv \frac{1}{\#\left(S^{te}\right)} \sum_{i \in S^{te}} \left\{\left(\tau_i - \hat{\tau}\left(X_i; S^{est}, \prod\right)\right)^2 - \tau_i^2\right\}$$

The adjusted expected mean square error for average treatment (EMSE) is the expectation of $MSE_\tau\left(S^{te}, S^{est}, \prod\right)$, over the test sample and the estimation sample

$$EMSE_\tau\left(\prod\right) \equiv \mathrm{E}_{S^{te}, S^{est}}\left[MSE\left(S^{te}, S^{est}, \prod\right)\right]$$

In the above exercise, the test sample and estimation sample are independent of each other, therefore known as honest estimation. Then the goal is to construct $\pi(\square)$ that maximizes the following honest criterion:

$$Q^H\left(\pi\right) = -\mathrm{E}_{S^{te}, S^{est}}\left[MSE_\tau\left(S^{te}, S^{est}, \pi\left(S^{tr}\right)\right)\right]$$

However, estimating MSE_τ is infeasible because we do not observe the τ_i. The work of Athey & Imbens 2015) shows a way to measure it. They expand the EMSE as:

$$-EMSE_\tau\left(\prod\right) = \mathrm{E}_{X_i}\left[\tau^2\left(X_i; \prod\right)\right] - \mathrm{E}_{S^{est}, X_i}\left[\mathrm{V}\left(\hat{\tau}^2\left(X_i; S^{est}, \prod\right)\right)\right]$$

when the component of the above expectation can be estimated using only the training sample, therefore leads to an estimator for the infeasible criterion that solely depends on and given as:

$$-EMSE_\tau\left(S^{tr}, N^{est}, \prod\right) = \frac{1}{N^{tr}} \sum_{i \in S^{tr}} \left(X_i; S^{tr}, \prod\right) - \left(\frac{1}{N^{tr}} + \frac{1}{N^{est}}\right) \cdot \sum_{\ell \in \prod} \left(\frac{s^2_{S^{tr}_{treat}}\left(\ell\right)}{p} + \frac{s^2_{S^{tr}_{control}}\left(\ell\right)}{1-p}\right)$$

which provides a causal tree, where $s^2_{S^{tr}_{treat}}\left(\ell\right)$ is the within-leaf variance on outcomes Y for $S^{tr}_{control}$ in leaf ℓ; $s^2_{S^{tr}_{control}}\left(\ell\right)$ is the counterpart for S^{tr}_{treat}, and $p = N_{treat}/N$ is the treatment probability. Rearrangement of above equation gives:

$$-EMSE_{\tau}\left(S^{tr},N^{est},\Pi\right)=\alpha\frac{1}{N^{tr}}\sum_{i\in S^{tr}}\left(X_i;S^{tr},\Pi,\alpha\right)-\left(1-\alpha\right)\left(\frac{1}{N^{tr}}+\frac{1}{N^{est}}\right)\cdot\sum_{\ell\in\Pi}\left(\frac{s^2_{S^{tr}_{treat}}\left(\ell\right)}{p}+\frac{s^2_{S^{tr}_{control}}\left(\ell\right)}{1-p}\right)$$

where $\alpha\in\left(0,1\right)$ is a parameter to adjust the portion of MSE and the variance of $EMSE$.

6. Estimation of treatment policies

Once, policymaker understand the heterogeneity effect, she would like to assign the correct treatment to each individual or subpopulation. For that, I implement Athey & Wager (2017) strategy to find the policy function π that can map the observable characteristic of individuals X_i to an available set of treatment W_i.

$$\pi:X_i\rightarrow W_i\in\left\{+1,-1\right\}$$

Note, here I redefine $W_i\in\left\{1,0\right\}$ to $W_i\in\left\{+1,-1\right\}$ which will help to formulate optimal policy assignment strategy later. Then an optimal treatment assignment policy can be given as π^* that maximizes expected utility, in our case, the health outcomes.

$$\pi^*\in\arg\max_{\pi\in\prod}E\left[Y_i\left(\pi\left(X_i\right)\right)\right]$$

Alternatively, any other non-optimal policy experiences the regret of $R\left(\pi\right)$, and we would like to minimize the regret function:

$$R\left(\pi\right)=E\left[Y_i\left(\pi^*\left(X_i\right)\right)\right]-E\left[Y_i\left(\pi\left(X_i\right)\right)\right]$$

Under unconfoundedness and the overlap assumption and binary treatment assignment Athey & Wager (2017) purpose a technique to estimate the regret, regret convergence and bound of the regret. They first determine the treatment effect $\hat{\Gamma}_i$ for each i using the double robust estimation technique called double machine learning of Chernozhukov et al. (2018) and given as:

$$\hat{\Gamma}_i=\hat{\mu}^{-k(i)}_{+1}\left(X_i\right)-\hat{\mu}^{-k(i)}_{-1}\left(X_i\right)+W_i\frac{Y_i-\hat{\mu}^{-k(i)}_{W_i}\left(X_i\right)}{\hat{e}^{-k(i)}_{W_i}\left(X_i\right)}$$

This is doubly-robust estimator because only one of $\hat{\mu}$ or $\hat{e}$ needs to be correctly specified, and the term double machine learning is used because $\hat{\mu}$ and $\hat{e}$ can be semi- or non-parametric estimators. I use L_1-penalty logistic regression to estimate $\hat{\mu}$ and $\hat{e}$ with $k(i)$ fold cross-validation.

If estimate is a positive treatment effect $\hat{\Gamma}_i$ I assign individual to treatment $\left(\pi\left(X_i\right)=1\right)$ and if nt then I assign individual to control $\left(\pi\left(X_i\right)=0\right)$ and penalize for mismatch and maximize the following Q function to assess the effective policy:

$$\hat{Q}(\pi)=n^{-1}\sum_i \pi\left(X_i\right)\left|\hat{\Gamma}_i\right| sign\left(\hat{\Gamma}_i\right)$$

Further, Athey & Wager (2017) show that the regret has $\sqrt{n}\left(\hat{R}_{DML}(\pi)-R(\pi)\right)\xrightarrow{d}N\left(0,\sigma^2(\pi)\right)$ convergence and is bounded with the order of $\sqrt{VC(\Pi)/n}$ where $\hat{R}_{DML}(\pi)$ is the double machine learning estimates of regret. The bound provides a robust theoretical prediction that test error on any out of sample data is upper bounded with the sum of training error and $\sqrt{VC(\Pi)/n}$.

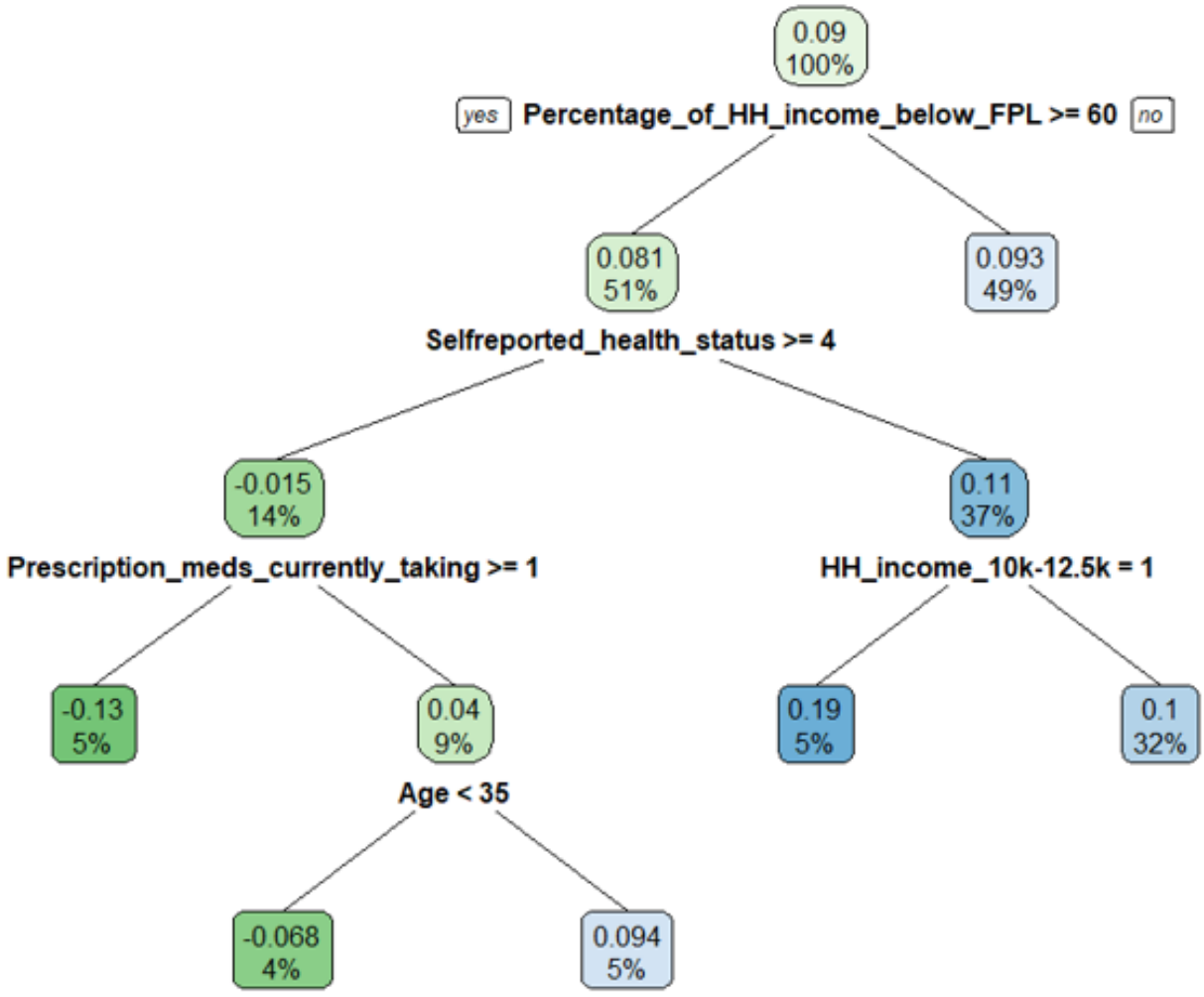
5. Results

In this paper, the data is randomly choosen in training (67%) and testing (33%) sample. Then CART is generated by applying the honest estimation as depicted in causal tree section. The training sample further splits randomly into three sets — the first trains, second estimates and third tests the CART model. The results section exhibits estimate for the causal tree which shows the magnitude of heterogeneous treatment effect. Then, it presents an optimal treatment policy design.

5.1 Causal Tree

Figure 2 shows an estimate of the treatment effect of having insurance among different subpopulation and Table 1 shows the training, estimation and testing sample OLS regression of the causal tree. Given the unconfoundedness and overlapping assumptions hold, the simple difference in each terminal node is likely to represent the causal effect.

Figure-2: Heterogeneous treatment effect with the causal tree diagram



Consider the subpopulation at baseline with household income below the federal poverty line by more than 60 percent and with self-reported health just above 4 and are taking more than one prescription medicines — after exposure to a year of insurance, the average self-reported health is on average 0.31 units lesser than control individual in this subgroup. The average health status of the treatment group is 3.288 in training sample, 3.189 in the estimation sample and 3.166 in the testing sample. These coefficients are statistically significant in 1% level of significance. This subgroup in the baseline is healthy individuals, once individual receive insurance by lottery, it seems individuals in the treatment group become attentive toward their health outcome. Thus, it conceivably shows habit biases in health outcome reporting. The data comprises about five percent of this subgroup.

Consider another subpopulation at baseline below 35 age with household income below the federal poverty line by more than 60 percent whose self-reported health is above 4 and are taking one or no prescription medicines. The average self-reported health is 0.07 units lower than the control group after exposure to a year of insurance. This result also hints possible habit bias among the younger subpopulation. The mean health outcome of treatment group is 3.658, 3.523 and 3.642 in training, estimation and testing sample respectively. These estimates are statistically significant in 1% level of significance. The data comprises about four percent of this subgroup. Again, examine this subpopulation stratum only with age of above 35, after exposure to a year of insurance, the average self-reported health is 0.094 units higher self-reported health than the control group. The mean health outcome of treatment group is about 2.512, 2.542 and 2.518 in training, estimation and testing sample. These estimates are statistically significant in 1% level of significance. The data comprises about five percent of this subgroup.

Consider another subpopulation at baseline with household income below the federal poverty line by more than 60 percent and with self-reported health 4 or below. If the baseline household income is between 10,000 to 12,500 dollars, after a year of insurance exposure, the treatment group self-reported health average is 0.19 unit higher than the control group. The average health status of treated in the training, estimation and testing example is 2.551, 2.564, 2.579 units respectively and statistically significant in 1% level of significance. This subpopulation is about five percent of the sample. While the rest of the subpopulation other than within 10,000 to 12,500 dollars household income bracket, after a year of insurance exposure, the treatment group self-reported health average is 0.1 units higher than the control group. This subpopulation comprises about 32 percent of the sample.

Table 1: Heterogeneous treatment effect of insurance on self-reported health

	Self reported health outcomes after a year		
	Training	Estimation	Testing
	(1)	(2)	(3)
leaf -0.1309	3.288***	3.189***	3.166***
	(0.062)	(0.064)	(0.067)
leaf -0.068	3.658***	3.523***	3.642***
	(0.070)	(0.078)	(0.075)
leaf 0.093	2.512***	2.542***	2.518***
	(0.020)	(0.020)	(0.020)
leaf 0.0944	3.195***	3.206***	3.109***
	(0.066)	(0.062)	(0.062)
leaf 0.1024	2.551***	2.564***	2.579***
	(0.026)	(0.026)	(0.026)
leaf 0.1865	2.553***	2.404***	2.456***
	(0.061)	(0.065)	(0.063)
leaf -0.1309:W	−0.301***	−0.131	−0.129
	(0.088)	(0.090)	(0.091)
leaf -0.068:W	−0.261***	−0.068	−0.067
	(0.098)	(0.103)	(0.102)
leaf 0.093:W	0.167***	0.093***	0.117***
	(0.029)	(0.029)	(0.029)
leaf 0.0944:W	0.172*	0.094	0.232***
	(0.091)	(0.088)	(0.087)
leaf 0.1024:W	0.114***	0.102***	0.103***
	(0.037)	(0.036)	(0.037)
leaf 0.1865:W	−0.190**	0.186**	0.132
	(0.088)	(0.093)	(0.091)
Observations	8,808	8,807	8,808
R^2	0.890	0.891	0.891
Adjusted R^2	0.890	0.891	0.890
F Statistic	5,954.098***	5,973.822***	5,965.267***
Note:		*p<0.1; **p<0.05; ***p<0.01	

5.2 Estimation of treatment policies

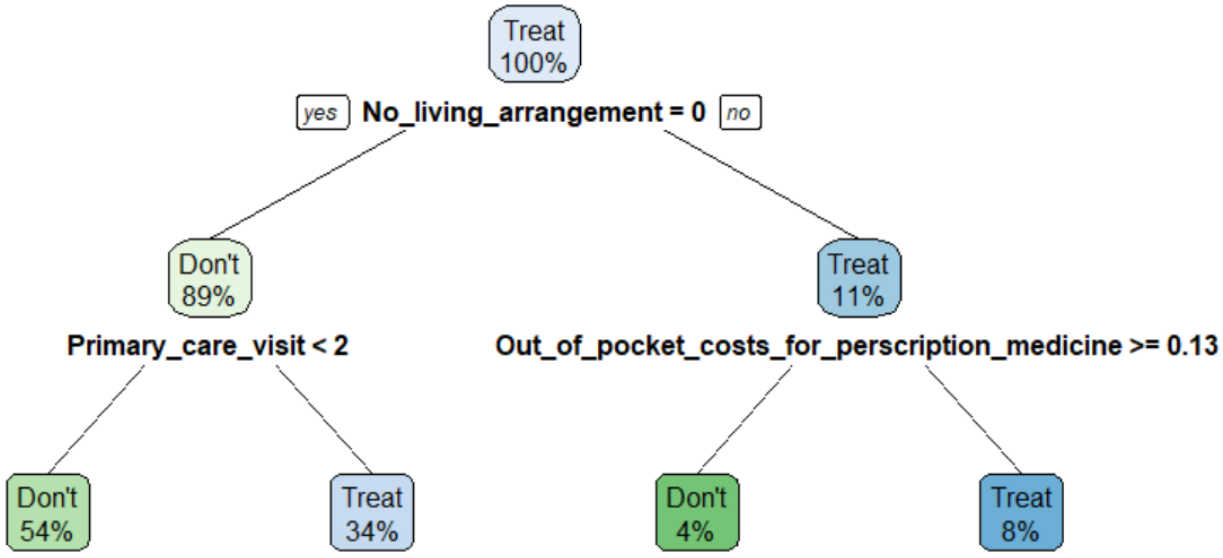
The primary reason that we would like to understand the heterogeneous treatment effect is to learn the correct treatment assignment to the subpopulation. In this section, I assume that policymaker would like to see if an individual with insurance reports higher health status or not after an intervention. Therefore, I take the difference in self-reported health status from the annual

survey to the baseline survey. If the difference is positive, I assign 1 to indicate that these individuals stated higher self-health regardless of treatment assignment and 0 otherwise. Further, in this section, I assign positive 1 to treated individual and negative 1 or else otherwise $W_i \in \{+1, -1\}$.

I estimate the treatment effect with double machine learning which is a doubly robust estimator. I use L_1-penalty logistic regression with to estimate $\hat{\mu}$ and $\hat{e}$ with five-fold cross-validation. The $\hat{\mu}$ and $\hat{e}$ provides an estimation of the Q function. Maximizing the Q function by choosing policy design yields an optimal policy. Figure 3 exhibits the optimal policy design.

Figure 3 displays two groups which minimize the regret of insurance assignment policy. The first target group is the subpopulation with no living arrangement who visit more than two primary care in the past six months. The second target group is subpopulation who are paying for their prescription medicine from out of pocket and do have some living arrangement. Insurance assignment to these two subpopulations maximizes the overall self-reported health.

Figure-3: Optimal policy



6. Conclusion

This paper estimates the causal impact of insurance on health. Seems obvious, but due to the endogeneity which arises through selection and heterogeneity treatment effect contaminates the estimates. Theoretically, a random assignment of treatment is not prone to endogeneity. This

paper uses the lottery assignment of insurance. The data of treatment assignment, health status outcome and observable characteristics are obtained from the Oregon health insurance experiment. Before estimating the results, this paper documents the identification assumption of unconfoundedness and overlapping.

Several studies exist that present the average treatment effect of health insurance using this Oregon health insurance experiment. However, this paper provides a first hand and novel results that show the existence of heterogeneity treatment effect. The knowledge of heterogeneous treatment effect is essential to understand the subpopulation that responds differently for the same treatment assignment. This paper contributes to three separate results — first, the plausibility of healthy habit bias among healthy poor; second, the plausibility of healthy habit bias among healthy and younger poor and third, the subpopulation with the highest turnover in health status response after a year of insurance exposure.

After revealing the presence of heterogeneous treatment effect, this paper attempts to learn the optimal policy. The optimal policy represents an intervention strategy that can maximize the overall increment of self-reported health. Under the assumption that the policymaker would like to maximize the total health status shifts, this paper design the policy that minimizes the policy regret. The optimal strategy is defined for two subpopulations. The first target group is the subpopulation with no living arrangement who visit more than two primary care in the past six months. The second target group is subpopulation who are paying for their prescription medicine from out of pocket and do have some living arrangement. Insurance assignment to these two subpopulations maximizes the overall self-reported health.

References

Allen, H., Baicker, K., Finkelstein, A., Taubman, S., & Wright, B. J. (2010). What the Oregon Health Study can Tell us about Expanding Medicaid. *Health Affairs*, *29*(8), 1498–1506. https://doi.org/10.1377/hlthaff.2010.0191

Assmann, S. F., Pocock, S. J., Enos, L. E., & Kasten, L. E. (2000). Subgroup analysis and other (mis)uses of baseline data in clinical trials. *The Lancet*, *355*(9209), 1064–1069. https://doi.org/10.1016/S0140-6736(00)02039-0

Athey, S. (2018). The Impact of Machine Learning on Economics. In *The Economics of Artificial Intelligence: An Agenda* (pp. 1–27). University of Chicago Press. https://doi.org/10.1257/jep.31.2.87

Athey, S., & Imbens, G. (2015). Recursive Partitioning for Heterogeneous Causal Effects, (October 2013), 1–22. https://doi.org/10.1073/pnas.1510489113

Athey, S., & Imbens, G. W. (2017). The Econometrics of Randomized Experiments. *Handbook of Economic Field Experiments*, *1*, 73–140. https://doi.org/10.1016/BS.HEFE.2016.10.003

Athey, S., & Wager, S. (2017). Efficient Policy Learning.

Athey, S., & Wager, S. (2018). Efficient Policy Learning, 1–37. https://doi.org/10.1007/978-3-642-57410-8_4

Atkins, J. P., Burdon, D., & Allen, J. H. (2007). An application of contingent valuation and decision tree analysis to water quality improvements. *Marine Pollution Bulletin*, *55*(10–12), 591–602. https://doi.org/10.1016/j.marpolbul.2007.09.018

Baicker, K., Allen, H. L., Wright, B. J., & Finkelstein, A. N. (2017). The Effect Of Medicaid On Medication Use Among Poor Adults: Evidence from Oregon. *Health Affairs*, *36*(12), 2110–2114. https://doi.org/10.1377/hlthaff.2017.0925

Baicker, K., & Finkelstein, A. (2011). The Effects of Medicaid Coverage — Learning from the Oregon Experiment. *New England Journal of Medicine*, *365*(8), 683–685. https://doi.org/10.1056/NEJMp1108222

Baicker, K., Finkelstein, A., Song, J., & Taubman, S. (2014). The Impact of Medicaid on Labor Market Activity and Program Participation: Evidence from the Oregon Health Insurance Experiment. *American Economic Review*, *104*(5), 322–328. https://doi.org/10.1257/aer.104.5.322

Baicker, K., Taubman, S. L., Allen, H. L., Bernstein, M., Gruber, J. H., Newhouse, J. P., … Finkelstein, A. N. (2013). The Oregon Experiment — Effects of Medicaid on Clinical Outcomes. *New England Journal of Medicine*, *368*(18), 1713–1722. https://doi.org/10.1056/NEJMsa1212321

Ballotpedia. (2018). Oregon Measure 101, Healthcare Insurance Premiums Tax for Medicaid Referendum (January 2018) - Ballotpedia. Retrieved May 16, 2019, from https://ballotpedia.org/Oregon_Measure_101,_Healthcare_Insurance_Premiums_Tax_for_Medicaid_Referendum_(January_2018)

Breiman, L. (2001). Random Forrests. *Machine Learning*.

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). Classification and regression trees Regression trees. *Wadsworth: Belmont, CA*, (June), 358. https://doi.org/10.1002/widm.8

Brook, R. H., Ware, J. E., Rogers, W. H., Keeler, E. B., Davies, A. R., Donald, C. A., … Newhouse, J. P. (1983). Does Free Care Improve Adults' Health? *New England Journal of Medicine*. https://doi.org/10.1056/NEJM198312083092305

Card, D., & Maestas, N. (2008). Care Utilization : Evidence from Medicare. *American Economic Review*, *98*(5), 2242–2258. https://doi.org/10.1257/aer.98.5.2242.The

Chen, Y. L., Hu, H. W., & Tang, K. (2009). Constructing a decision tree from data with hierarchical class labels. *Expert Systems with Applications*, *36*(3 PART 1), 4838–4847. https://doi.org/10.1016/j.eswa.2008.05.044

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *Econometrics Journal*, *21*(1), C1–C68. https://doi.org/10.1111/ectj.12097

Chernozhukov, V., Escanciano, J. C., Newey, W. K., & Robins, J. M. (2018). Locally Robust Semiparametric Estimation arXiv : 1608 . 00033v2 [math . ST] 31 May 2018, 1–48.

Cook, D., Gebski, V., & Keech, A. (2004). Subgroup analysis in clinical trials. *Med J Aust*, *180*(March), 289–291.

Crump, R. K., Hotz, V. J., Imbens, G. W., & Mitnik, O. A. (2008). Nonparametric Tests for Treatment Effect Heterogeneity. *The Review of Economics and Statistics*, *90*(3), 389–405. https://doi.org/10.1162/rest.90.3.389

Currie, J., & Gruber, J. (1996a). Health Insurance Eligibility, Utilization of Medical Care, and Child Health. *The Quarterly Journal of Economics*, *111*(2), 431–466. https://doi.org/10.2307/2946684

Currie, J., & Gruber, J. (1996b). Saving Babies : The Efficacy and Cost of Recent Changes in the Medicaid Eligibility of Pregnant Women. *Journal of Political Economy*, *104*(6), 1263–1296.

Davis, J. M. V., & Heller, S. B. (2017). Using Causal Forests to Predict Treatment Heterogeneity: An application to

Summer Jobs. *American Economic Review*, *107*(5), 546–550. https://doi.org/10.1257/aer.p20171000

Finkelstein, A., & McKnight, R. (2008). What Did Medicare Do? The Initial Impact of Medicare on Mortality and Out of Pocket Medical Spending. *Journal of Public Economics*. https://doi.org/10.1016/j.jpubeco.2007.10.005

Finkelstein, A., Taubman, S., Wright, B., Bernstein, M., Gruber, J., Newhouse, J. P., … Oregon Health Study Group, . (2012). The Oregon Health Insurance Experiment: Evidence From The First Year. *Quarterly Journal of Economics*, *127*(August (3)), 1057–1106. https://doi.org/10.1093/qje/qjs020.Advance

Foden-Vencil, K. (2018). Oregon Measure 101: What You Need To Know . News | OPB. Retrieved May 16, 2019, from https://www.opb.org/news/article/oregon-measure-101-faq-medicaid-tax-ballot-health-care/

Grossman, R. L., Heath, A. P., Ferretti, V., Varmus, H. E., Lowy, D. R., Kibbe, W. A., & Staudt, L. M. (2016). Effect of Medicaid Coverage on ED Use — Further Evidence from Oregon's Experiment. *New England Journal of Medicine*, *363*(1), 1–3. https://doi.org/10.1056/NEJMp1002530

Hanratty, M. J. (1996). American Economic Association Canadian National Health Insurance and Infant Health. *The American Economic Review*, *86*(1), 276–284.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*. https://doi.org/10.1080/01621459.1986.10478354

Horner, S. B., Fireman, G. D., & Wang, E. W. (2010). The relation of student behavior, peer status, race, and gender to decisions about school discipline using CHAID decision trees and regression modeling. *Journal of School Psychology*, *48*(2), 135–161. https://doi.org/10.1016/j.jsp.2009.12.001

Imbens, G. W., & Rubin, D. B. (2015a). *Causal inference: For statistics, social, and biomedical sciences an introduction*. *Causal Inference: For Statistics, Social, and Biomedical Sciences an Introduction*. https://doi.org/10.1017/CBO9781139025751

Imbens, G. W., & Rubin, D. B. (2015b). *Causal inference: For statistics, social, and biomedical sciences an introduction*. *Causal Inference: For Statistics, Social, and Biomedical Sciences an Introduction*. https://doi.org/10.1017/CBO9781139025751

Kim, H., & Koehler, G. J. (1995). Theory and practice of decision tree induction. *Omega*, *23*(6), 637–652. https://doi.org/10.1016/0305-0483(95)00036-4

Kitagawa, T., & Tetenov, A. (2018). Who Should Be Treated? Empirical Welfare Maximization Methods for Treatment Choice. *Econometrica*. https://doi.org/10.3982/ecta13288

Kleinberg, J., Ludwig, J., Mullainathan, S., & Obermeyer, Z. (2015). Prediction Policy Problems. *American Economic Review*. https://doi.org/10.1257/aer.p20151023

Lee, M. J. (2009). Non-parametric tests for distributional treatment effect for randomly censored responses. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *71*(1), 243–264. https://doi.org/10.1111/j.1467-9868.2008.00683.x

Levy, H., & Meltzer, D. (2004). What Do We Really Know About Whether Health Insurance Affects Health? *Health Policy and the Uninsured*, 179–204.

Levy, H., & Meltzer, D. (2008). The Impact of Health Insurance on Health. *Annual Review of Public Health*, *29*(1), 399–409. https://doi.org/10.1146/annurev.publhealth.28.021406.144042

McWilliams, J. M., Meara, E., Zaslavsky, A. M., & Ayanian, J. Z. (2007a). Health of Previously Uninsured Adults after Acquiring Medicare Coverage. *JAMA - Journal of the American Medical Association*. https://doi.org/10.1001/jama.298.24.2886

McWilliams, J. M., Meara, E., Zaslavsky, A. M., & Ayanian, J. Z. (2007b). Use of Health Services by Previously Uninsured Medicare Beneficiaries. *New England Journal of Medicine*. https://doi.org/10.1056/NEJMsa067712

Mullainathan, S., & Spiess, J. (2017). Machine Learning: An Applied Econometric Approach. *Journal of Economic*

Perspectives—Volume, *31*(2—Spring), 87–106. https://doi.org/10.1257/jep.31.2.87

Newhouse, J. P. (1994). Free for All: Lessons from the RAND Health Insurance Experiment. *BMJ*. https://doi.org/10.1136/bmj.308.6945.1724a

Norris, L. (2018). Oregon and the ACA's Medicaid expansion: eligibility, enrollment and benefits | healthinsurance.org. Retrieved May 16, 2019, from https://www.healthinsurance.org/oregon-medicaid/#enrollment

Oh, J., & Kim, B. (2010). Prediction Model for Demands of the Health Meteorological Information Using a Decision Tree Method. *Asian Nursing Research*, *4*(3), 151–162. https://doi.org/10.1016/S1976-1317(10)60015-1

Pearl, J. (2011). *Causality: Models, Reasoning, and Inference, Second Edition*. *Causality: Models, Reasoning, and Inference, Second Edition*. https://doi.org/10.1017/CBO9780511803161

Rosenbaum, P. R., & Rubin, D. B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, *70*(1), 41–55. https://doi.org/10.1093/biomet/70.1.41

Rubin, D. B. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*. https://doi.org/10.1037/h0037350

Rubin, D. B. (1978). Bayesian Inference for Causal Effects: The Role of Randomization. *The Annals of Statistics*, *6*(1), 34–58. https://doi.org/10.1214/aos/1176344064

Shakya, S., & Lama, N. (2014). Possible Decision Rules to Allocate Quotas and Reservations to Ensure Equity for Nepalese Poor. *Economic Journal of Development Issues*, *17*(1), 149–162.

Taubman, S. L., Allen, H. L., Wright, B. J., & Baicker, K. (2014). Oregon's Health Insurance Experiment. *Science*, *343*(6168), 263–268. https://doi.org/10.1126/science.1246183

Varian, H. R. (2014). Big Data: New Tricks for Econometrics. *Journal of Economic Perspectives*, *28*(2), 3–28. https://doi.org/10.1257/jep.28.2.3

Wager, S., & Athey, S. (2017). Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. https://doi.org/10.1073/pnas.1510489113

Willke, R. J., Zheng, Z., Subedi, P., Althin, R., & Mullins, C. D. (2012). From concepts, theory, and evidence of heterogeneity of treatment effects to methodological approaches: A primer. *BMC Medical Research Methodology*. https://doi.org/10.1186/1471-2288-12-185

Zhou, R. A., Baicker, K., Taubman, S., & Finkelstein, A. N. (2017). The uninsured do not use the emergency department more- they use other care less. *Health Affairs*, *36*(12), 2115–2122. https://doi.org/10.1377/hlthaff.2017.0218

Annex

Table 1: Annex

Variables	Null Weighted			
	Control	Treated	SMD	p
n	13152	13271		
Age	42.16 (12.18)	41.68 (12.16)	0.04	0.001
Prescription medicine	0.98 (1.45)	0.98 (1.44)	0.002	0.878
Primary care visits	1.37 (1.62)	1.36 (1.61)	0.005	0.661
Number of ER visits, truncated at 2*99th percentile	0.30 (0.60)	0.30 (0.59)	0.006	0.631
Prescription medicine Out of pocket	20.69 (27.29)	20.07 (27.09)	0.023	0.066
Any out of pocket costs for medical care in the past 6 months	0.68 (0.47)	0.66 (0.47)	0.044	0.001
Total amount currently owed for medical expenses	373.71 (700.73)	360.57 (694.98)	0.019	0.126
Self-reported health status at baseline	2.77 (1.08)	2.82 (1.07)	0.048	0.001
Household Size (adults and children)	2.81 (1.49)	2.91 (1.50)	0.071	0.001
Household income as percent of federal poverty line	62.18 (54.52)	65.07 (55.22)	0.053	0.001
Number of family members under 19 living in house	0.80 (1.17)	0.88 (1.20)	0.068	0.001
Number of family members under age 19 living in house: insurance status	2.21 (0.95)	2.13 (0.97)	0.078	0.001
Total out of pocket costs for medical care, last 6 months (corrected version)	292.38 (410.62)	285.75 (410.23)	0.016	0.189
In the last 6 months, how much money did you spend on medical care?	113.58 (156.42)	111.44 (156.53)	0.014	0.266

Table 2: Annex

Variables	Weighted with IPW			
	Control	Treated	SMD	p
n	26425.89	26422.34		
Age	41.89 (12.18)	41.90 (12.19)	0.001	0.936
Prescription medicine	0.98 (1.45)	0.98 (1.43)	0.001	1
Primary care visits	1.36 (1.62)	1.36 (1.61)	0.001	0.987
Number of ER visits, truncated at 2*99th percentile	0.30 (0.60)	0.30 (0.60)	0.001	0.943
Prescription medicine Out of pocket	20.41 (27.02)	20.41 (27.30)	0.001	0.997
Any out of pocket costs for medical care in the past 6 months	0.67 (0.47)	0.67 (0.47)	0.001	0.979
Total amount currently owed for medical expenses	367.19 (692.87)	366.98 (700.77)	0.001	0.981
Selfreported health status at baseline	2.80 (1.08)	2.80 (1.07)	0.001	0.982
Household Size (adults and children)	2.86 (1.50)	2.86 (1.49)	0.001	0.948
Household income as percent of federal poverty line	64.07 (55.03)	64.02 (55.01)	0.001	0.939
Number of family members under 19 living in house	0.85 (1.20)	0.85 (1.18)	0.001	0.947
Number of family members under age 19 living in house: insurance status	2.17 (0.96)	2.17 (0.96)	0.001	0.999
Total out of pocket costs for medical care, last 6 months (corrected version)	289.32 (411.55)	289.58 (410.18)	0.001	0.96
In the last 6 months, how much money did you spend on medical care?	113.11 (156.27)	113.06 (157.32)	0.001	0.978

Table 3: Annex

Variables	Null Weighted			
	Control	Treated	SMD	p
n	13152	13271		
Gave a phone number on lottery sign up	11801.0 (89.7)	11952.0 (90.1)	0.011	0.369
Speaks english	12246.0 (93.1)	12207.0 (92.0)	0.043	0.001
Female	7872.0 (59.9)	7740.0 (58.3)	0.031	0.011
Signed up for lottery list on first day	1344.0 (10.2)	1400.0 (10.5)	0.011	0.379
Signed up for lottery list on last day	464.0 (3.5)	432.0 (3.3)	0.015	0.221
Gave a PO Box as an address	1697.0 (12.9)	1654.0 (12.5)	0.013	0.283
Individual signed him or herself up for the lottery list	11627.0 (88.4)	11208.0 (84.5)	0.116	0.001
Zip code from lottery list is a metropolitan statistical area	9834.0 (74.8)	9914.0 (74.7)	0.002	0.899
Self-describe as: Spanish, Hispanic or Latino	1352.0 (10.3)	1529.0 (11.5)	0.04	0.001
Self-identify as: white	10899.0 (82.9)	10819.0 (81.5)	0.035	0.004
Self-identify as: black	474.0 (3.6)	467.0 (3.5)	0.005	0.709
Self-identify as: American Indian or Alaska Native	809.0 (6.2)	725.0 (5.5)	0.029	0.017
Self-identify as: asian	387.0 (2.9)	445.0 (3.4)	0.024	0.056
Self-identify as: native hawaiian or pacific islander	119.0 (0.9)	128.0 (1.0)	0.006	0.614
Self-identify as: other race	1093.0 (8.3)	1145.0 (8.6)	0.011	0.354
Currently have OHP insurance	8785.0 (66.8)	7060.0 (53.2)	0.28	0.001
Needed medical care in the last six months	10222.0 (77.7)	10030.0 (75.6)	0.051	0.001
Needed prescription medications in the last six months	9376.0 (71.3)	9197.0 (69.3)	0.044	0.001
Needed dental care in the last six months	10106.0 (76.8)	9901.0 (74.6)	0.052	0.001
Borrowed money/skipped bills to pay health care bills in last 6 months?	6037.0 (45.9)	5818.0 (43.8)	0.041	0.001
Diagnosed with Diabetes	1503.0 (11.4)	1389.0 (10.5)	0.031	0.012
Diagnosed with Asthma	2180.0 (16.6)	2101.0 (15.8)	0.02	0.101
Diagnosed with High blood pressure	3740.0 (28.4)	3549.0 (26.7)	0.038	0.002
Diagnosed with COPD	1010.0 (7.7)	960.0 (7.2)	0.017	0.168
Diagnosed with Congestive Heart Failure	339.0 (2.6)	332.0 (2.5)	0.005	0.695
Diagnosed with Depression or Anxiety	5844.0 (44.4)	5455.0 (41.1)	0.067	0.001

Table 4: Annex

Variables	Weighted with IPW			
	Control	Treated	SMD	p
n	26426	26422		
Gave a phone number on lottery sign up	23754.2 (89.9)	23752.9 (89.9)	0.001	0.985
Speaks english	24432.5 (92.5)	24437.1 (92.5)	0.001	0.928
Female	15629.7 (59.1)	15625.5 (59.1)	0.001	0.99
Signed up for lottery list on first day	2735.6 (10.4)	2735.0 (10.4)	0.001	0.999
Signed up for lottery list on last day	903.8 (3.4)	904.3 (3.4)	0.001	0.992
Gave a PO Box as an address	3341.6 (12.6)	3338.7 (12.6)	0.001	0.983
Individual signed him or herself up for the lottery list	22823.2 (86.4)	22828.9 (86.4)	0.001	0.939
Zip code from lottery list is a metropolitan statistical area	19768.8 (74.8)	19770.2 (74.8)	0.001	0.978
Self-describe as: Spanish, Hispanic or Latino	2908.6 (11.0)	2899.8 (11.0)	0.001	0.936
Self-identify as: white	21672.9 (82.0)	21679.7 (82.1)	0.001	0.939
Self-identify as: black	941.9 (3.6)	942.3 (3.6)	0.001	0.993
Self-identify as: American Indian or Alaska Native	1538.7 (5.8)	1541.1 (5.8)	0.001	0.973
Self-identify as: asian	849.6 (3.2)	846.1 (3.2)	0.001	0.954
Self-identify as: native hawaiian or pacific islander	249.9 (0.9)	249.0 (0.9)	0.001	0.979
Self-identify as: other race	2261.6 (8.6)	2257.2 (8.5)	0.001	0.965
Currently have OHP insurance	15840.8 (59.9)	15842.8 (60.0)	0.001	0.98
Needed medical care in the last six months	20211.6 (76.5)	20220.8 (76.5)	0.001	0.933
Needed prescription medications in the last six months	18532.8 (70.1)	18548.5 (70.2)	0.002	0.904
Needed dental care in the last six months	19983.5 (75.6)	19993.1 (75.7)	0.001	0.931
Borrowed money/skipped bills to pay health care bills in last 6 months?	11860.5 (44.9)	11861.1 (44.9)	0.001	0.989
Diagnosed with Diabetes	2907.6 (11.0)	2907.8 (11.0)	0.001	0.995
Diagnosed with Asthma	4269.8 (16.2)	4271.3 (16.2)	0.001	0.986
Diagnosed with High blood pressure	7283.4 (27.6)	7276.2 (27.5)	0.001	0.967
Diagnosed with COPD	1959.6 (7.4)	1962.9 (7.4)	0.001	0.967
Diagnosed with Congestive Heart Failure	670.2 (2.5)	669.6 (2.5)	0.001	0.992
Diagnosed with Depression or Anxiety	11278.1 (42.7)	11275.5 (42.7)	0.001	0.995

Table 5: Annex

Variables	Null Weighted			
	Control	Treated	SMD	p
Currently employed or self-employed			1	0.001
Yes, employed	4824.0 (36.7)	5041.0 (38.0)		
Yes, self-employed	1472.0 (11.2)	1416.0 (10.7)		
not employed	6398.0 (48.6)	6391.0 (48.2)		
retired	458.0 (3.5)	423.0 (3.2)		
Average hrs worked per week				
don't currently work	6930.0 (52.7)	6899.0 (52.0)		
work less 20 hrs per week	1242.0 (9.4)	1235.0 (9.3)		
work 20-29 hrs per week	1422.0 (10.8)	1431.0 (10.8)		
work 30+ hrs perweek	3558.0 (27.1)	3706.0 (27.9)		
Living arrangement			0.092	0.001
Live alone	2329.0 (17.7)	2060.0 (15.5)		
Live with partner or spouse	5452.0 (41.5)	6048.0 (45.6)		
Live with parents	1296.0 (9.9)	1193.0 (9.0)		
Live with other relatives	1960.0 (14.9)	1956.0 (14.7)		
Live with friends	1074.0 (8.2)	973.0 (7.3)		
Other	1041.0 (7.9)	1041.0 (7.8)		
Highest level of education completed			0.025	0.263
less than high school	2339.0 (17.8)	2467.0 (18.6)		
high school diploma or GED	6821.0 (51.9)	6816.0 (51.4)		
vocational or 2-year degree	2680.0 (20.4)	2633.0 (19.8)		
4-year degree	1312.0 (10.0)	1355.0 (10.2)		

Table 6: Annex

Variables	Weighted with IPW			
	Control	Treated	SMD	p
Currently employed or self-employed			0.081	0.032
Yes, employed	9905.9 (37.5)	9902.8 (37.5)		
Yes, self-employed	2891.5 (10.9)	2889.8 (10.9)		
not employed	12738.2 (48.2)	12737.0 (48.2)		
retired	890.2 (3.4)	892.8 (3.4)		
Average hrs worked per week				
don't currently work	13788.9 (52.2)	13789.6 (52.2)		
work less 20 hrs per week	2463.9 (9.3)	2464.0 (9.3)		
work 20-29 hrs per week	2852.2 (10.8)	2859.3 (10.8)		
work 30+ hrs perweek	7320.9 (27.7)	7309.5 (27.7)		
Living arrangement			0.001	1
Live alone	4406.2 (16.7)	4409.5 (16.7)		
Live with partner or spouse	11532.7 (43.6)	11525.4 (43.6)		
Live with parents	2468.8 (9.3)	2469.0 (9.3)		
Live with other relatives	3917.0 (14.8)	3913.6 (14.8)		
Live with friends	2038.2 (7.7)	2039.0 (7.7)		
Other	2063.0 (7.8)	2065.8 (7.8)		
Highest level of education completed			0.001	1
less than high school	4794.1 (18.1)	4789.6 (18.1)		
high school diploma or GED	13599.1 (51.5)	13605.7 (51.5)		
vocational or 2-year degree	5339.4 (20.2)	5336.9 (20.2)		
4-year degree	2693.2 (10.2)	2690.1 (10.2)		

Table 7: Annex

Variables	Null Weighted			
	Control	Treated	SMD	p
Household income category (in dollars)			0.109	0.001
No income	1880.0 (14.3)	1839.0 (13.9)		
1-2500	1707.0 (13.0)	1484.0 (11.2)		
2501-5000	1080.0 (8.2)	986.0 (7.4)		
5001-7500	940.0 (7.1)	968.0 (7.3)		
7501-10000	1155.0 (8.8)	1118.0 (8.4)		
10001-12500	1189.0 (9.0)	1196.0 (9.0)		
12501-15000	1145.0 (8.7)	1097.0 (8.3)		
15001-17500	731.0 (5.6)	781.0 (5.9)		
17501-20000	767.0 (5.8)	878.0 (6.6)		
20001-22500	510.0 (3.9)	599.0 (4.5)		
22501-25000	561.0 (4.3)	585.0 (4.4)		
25001-27500	318.0 (2.4)	329.0 (2.5)		
27501-30000	284.0 (2.2)	387.0 (2.9)		
30001-32500	226.0 (1.7)	232.0 (1.7)		
32501-35000	159.0 (1.2)	182.0 (1.4)		
35001-37500	87.0 (0.7)	125.0 (0.9)		
37501-40000	109.0 (0.8)	111.0 (0.8)		
40001-42500	71.0 (0.5)	78.0 (0.6)		
42501-45000	51.0 (0.4)	38.0 (0.3)		
45001-47500	31.0 (0.2)	43.0 (0.3)		
47501-50000	49.0 (0.4)	64.0 (0.5)		
more than 50000	102.0 (0.8)	151.0 (1.1)		

Table 8: Annex

Variables	Weighted with IPW			
	Control	Treated	SMD	p
Household income category (in dollars)			0.003	1
No income	3667.5 (13.9)	3676.0 (13.9)		
1-2500	3176.3 (12.0)	3169.5 (12.0)		
2501-5000	2059.2 (7.8)	2059.6 (7.8)		
5001-7500	1896.3 (7.2)	1898.7 (7.2)		
7501-10000	2266.4 (8.6)	2269.7 (8.6)		
10001-12500	2387.3 (9.0)	2390.3 (9.0)		
12501-15000	2247.1 (8.5)	2247.8 (8.5)		
15001-17500	1529.3 (5.8)	1529.2 (5.8)		
17501-20000	1656.8 (6.3)	1654.8 (6.3)		
20001-22500	1114.3 (4.2)	1113.2 (4.2)		
22501-25000	1162.6 (4.4)	1158.2 (4.4)		
25001-27500	658.6 (2.5)	656.2 (2.5)		
27501-30000	679.0 (2.6)	674.8 (2.6)		
30001-32500	459.9 (1.7)	461.3 (1.7)		
32501-35000	346.4 (1.3)	344.5 (1.3)		
35001-37500	209.6 (0.8)	210.8 (0.8)		
37501-40000	222.9 (0.8)	222.4 (0.8)		
40001-42500	151.2 (0.6)	150.6 (0.6)		
42501-45000	91.0 (0.3)	92.1 (0.3)		
45001-47500	72.3 (0.3)	73.1 (0.3)		
47501-50000	112.9 (0.4)	113.3 (0.4)		
more than 50000	258.9 (1.0)	256.3 (1.0)		