Heterogeneous Impact of Medicaid Expansion Program in Oregon and Generalizing Effects to Other States

Samin Jalali

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2022

Reading Committee:

Anirban Basu, Chair

Jing Tao, Chair

Yuya Takahashi

Program Authorized to Offer Degree:
Economics

University of Washington

Abstract

Heterogeneous Impact of Medicaid Expansion Program in Oregon and Generalizing Effects to Other States

Samin Jalali

Co-Chairs of the Supervisory Committee:
Anirban Basu
Department of Pharmacy

Jing Tao
Department of Economics

Medicaid is the largest means-tested program in the US. It is crucial to assess how it affects its subscribers. This dissertation provides new evidence of Medicaid's effect on several outcomes beyond physical health. Specifically, I study the heterogeneity of Medicaid's effect on self-reported happiness, self-reported depression, and self-reported out-of-pocket medical costs. Using Generalized Random Forest (GRF), a causal Machine Learning method, I estimate the effect of the Oregon's 2008 Medicaid expansion as a non-parametric function of individuals' characteristics.

I find that age, weekly working hours, and urbanicity create considerable heterogeneity in Medicaid's effects. My results show that the Medicaid coverage causes the older population who work more than 30 hours per week to be happier. It also causes adults living in rural areas to incur less out-of-pocket medical costs; and finally, young adults who work at least several hours per week to be less depressed.

My results also shed light on some of the mechanisms through which Medicaid may affect these outcomes. I show that the pent-up healthcare demand among older people, education-related efficient use of health services among young adults, and differential health care competition in urban and rural areas are plausible mechanisms which can explain heterogeneity in Medicaid's effects.

To examine what would be the effect of Medicaid beyond the Oregon population, I use the results from Oregon experiment to transport them to another State which has not implemented the Medicaid expansion program yet. I employ two statistical methods such as predicting effects using trained GRF on Oregon data and inverse propensity score weighting. Using these methods I estimate the average effect of Medicaid in the target population. My results can help policymakers to better understand the effect of Medicaid expansion and design targeted public insurance programs.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to express my gratitude to my advisors, Prof. Anirban Basu and Prof. Jing Tao for their invaluable guidance and support. The completion of this dissertation would not have been possible without their mentorship. I would also like to thank Prof. Yuya Takahashi and Prof. Bianca Frogner for kindly serving on my dissertation committee, Simon Reeve-Parker and Heidi Hannah from the UW Economics Department for providing extensive administrative support, my friends, Alborz, Abbas, Farnaz, Kamal, Reza and many others who were always there for me.

Above all, I am extremely thankful to my parents Hassan and Zohre, my sister and brother, Alieh and Amin for their endless support, encouragement, and motivation.

DEDICATION

To my family.

Chapter 1

INTRODUCTION

1.1 Background and Motivation

Medicaid is a joint program between Federal and State governments providing public health insurance for low-income individuals. Medicaid coverage expansion is the primary goal of the Affordable Care Act[1] (ACA) which gives the State the authority to implement the Medicaid expansion. As of February 2021, 39 States have accepted Federal funding to expand the Medicaid program under the ACA law, while 12 States have not.[2] Figure 1.1 shows the States which have and have not adopted the Medicaid expansion program.

Medicaid is the largest means-tested program in the United States covering more than 75 million individuals in 2019. It is also a large program in terms of the costs. In order to have a better understanding of its size, Figure 1.2 shows the Medicaid expenditure and the healthcare expenditure in the U.S. over years as a percentage of GDP. In 2019 about 3% of GDP was spent on Medicaid while the share of healthcare expenditure from GDP was about 18%.[3] Given the size of this program, it is important to assess how this program is affecting its subscribers.

Several studies have demonstrated the *average* causal effects of Medicaid expansion on the marginal patients affected by this policy on several outcomes (e.g., Finkelstein et al., 2012; Baicker et al., 2013; Taubman et al., 2014; Finkelstein et al., 2016; Allen et al., 2013; Baicker et al., 2017, 2014). Their results raise further questions about the mechanisms through which insurance may affect outcomes. In this study, I use Generalized Random Forest (GRF), a non-parametric causal Machine Learning method, developed by Athey et al. (2019), to unveil heterogeneity in average treatment effects of Medicaid expansion on

---

[1] The comprehensive health care reform law enacted in March 2010 (www.healthcare.gov)

[2] https://www.kff.org/medicaid

[3] https://usafacts.org/

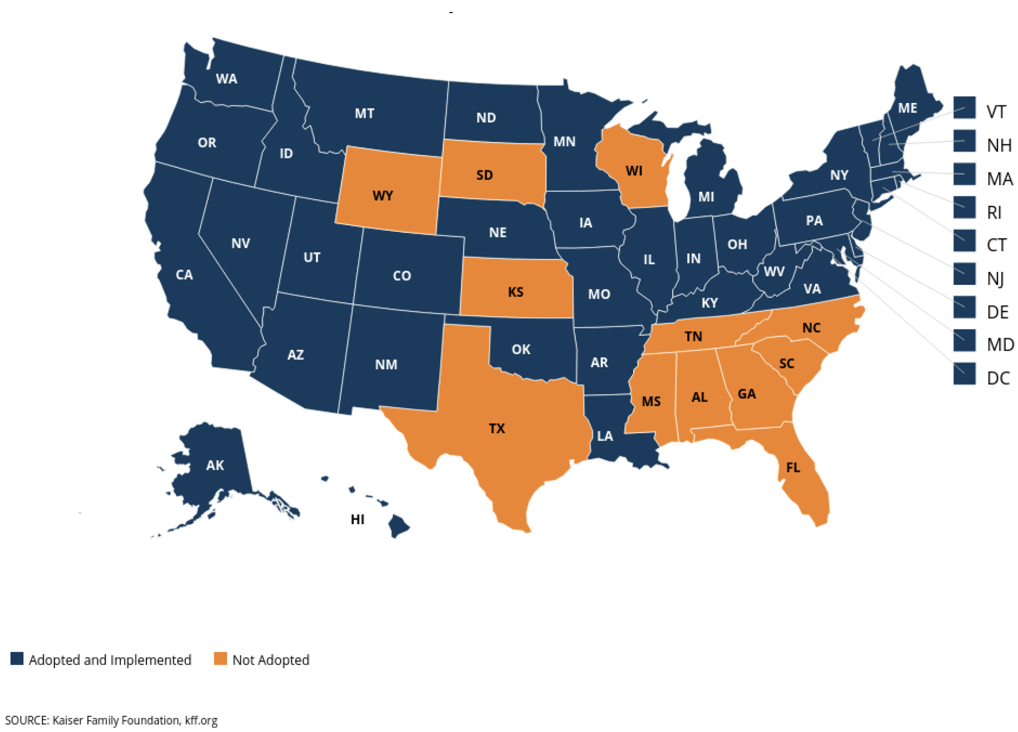

Figure (1.1) **Status of State Action on the Medicaid Expansion Decision**
This figure shows that, as of February 2021, a total of 39 states have already adopted the Medicaid expansion program and there are 12 states that have not yet adopted this program. The states shown in blue in this map are the ones that have adopted Medicaid and the orange ones have not.

various outcomes to better understand these mechanisms.

Estimating heterogeneous treatment effects in terms of observable characteristics of the subscribers allows us to evaluate Medicaid's effect at the individual level. This, in turn, can be used to find more about the subpopulations that benefit or are adversely affected the most from Medicaid expansion. These add more perspective to our understanding of the underlying mechanisms driving effects of Medicaid.

Health insurance is often considered a social determinant of health, although the effect of insurance can extend to outcomes beyond physical health. In this study, I provide new evidence for the impact of Medicaid expansion on self-reported happiness, self-reported out-of-pocket medical cost, and self-reported depression. Using the estimated heterogeneous effects I discuss mechanisms driving these effects.

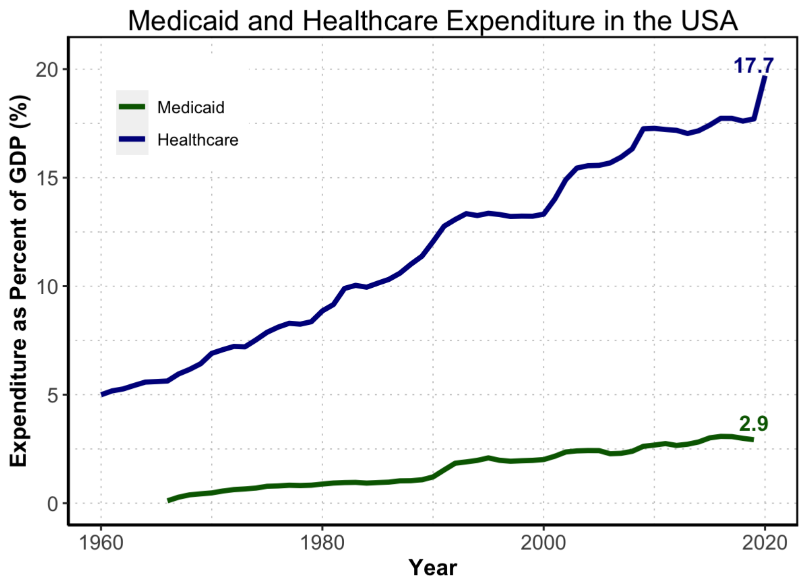

Figure (1.2) **Medicaid and Healthcare Expenditure as a Percentage of GDP** This plot shows the share of Medicaid and healthcare from GDP over years. In 2019 about 18% of GDP was spent on healthcare in the U.S. and about 3% of GDP was spent on Medicaid. These numbers show that Medicaid is a very large program.

1.2 Research Objectives

My analysis consists of four parts as explained below.

First, I replicate the estimated *average* causal effect of Medicaid on self-reported happiness, self-reported out-of-pocket medical cost, and self-reported depression as reported in the literature, specifically the results presented in Finkelstein et al. (2012). I reproduce these results in two ways: I use two-stage least squares which is the method used by Finkelstein et al. (2012), and as an alternative method, I aggregate the estimated individual treatment effect from GRF as a measure of average causal effect. My results from both methods show that Medicaid expansion increases happiness, decreases out-of-pocket medical costs, and decreases depression. These results are consistent with those reported in Finkelstein et al. (2012).

Second, I show that age is the most important covariate in creating heterogeneity in the effect of Medicaid for all of the selected outcomes. Additionally, weekly working hours and living in urban or rural area play important roles in explaining heterogeneity of the effect. I then show the variability of the estimated Medicaid's effect with respect to each of these variables. In the next step, I look at the heterogeneity of Medicaid's effect with respect to combinations of important variables (e.g., age and working hours) for each outcome. I show that Medicaid coverage causes the older population who work more than 30 hours per week to be happier. It also causes adults living in rural areas to incur less out-of-pocket medical costs; and finally, young adults who work at least several hours per week to be less depressed.

Third, having the heterogeneous treatment effect results at hand, I study competing mechanisms through which Medicaid may affect the selected outcomes. For explaining the heterogeneous effect of Medicaid on happiness I consider two potential competing mechanisms of pent-up health care demand and job lock. I find that the former is the more plausible mechanism for happiness. Similarly, I explore alternative mechanisms of baseline mental health burden versus education-related efficient use of health services to explain differential effects of Medicaid on depression in the younger population. My investigations show that the latter is the more plausible mechanism for depression. Correspondingly, by finding significant differences in baseline service-specific out-of-pocket medical cost between rural and urban areas, I propose that the differences in the effect of Medicaid on out-of-pocket medical costs could be driven by differential health care competition in urban areas.

Fourth, I use the estimated results from the previous parts of my study to generalize[4] to a different population which has not implemented the Medicaid expansion program.

1.3 Data and Methodology

I draw my data from Oregon Health Insurance Experiment (OHIE) conducted by Finkelstein and her team. This experiment is a randomized controlled trial based on a series of lottery drawings. The randomized nature of the experiment makes it an ideal dataset to study the causal relationship between Medicaid and the outcomes. The dataset consists of several surveys administered at different time intervals. My dataset combines the information

[4] In this context, transporting the results to a different population is generally used in place of generalization.

from these surveys and covers about 16,000 observations with information on demographic characteristics, participation status in the State programs, individuals' utilization of the healthcare, and the financial circumstances of the individuals.

Estimating the causal effect of Medicaid is challenging as observable and unobservable characteristics of the individuals and of the providers motivate or dissuade the participants to self-select into the program. Therefore, Medicaid enrollment is confounded by observable and unobservable characteristics, making it an endogenous variable. As such, we are facing with the selection bias problem (Levy and Meltzer, 2008). To address this issue, I use the lottery allocation as an Instrumental Variable.

The method I use in this study, Generalized Random Forest (GRF), is a non-parametric causal Machine Learning method which estimates treatment effect as a function of observable characteristics. GRF extends the idea of Random Forest developed by Breiman (2001). These algorithms are similar such that they are both based on building multiple trees with random sub-sampling of the data and random selection of the covariates, but different in the rules they use to split the covariate space. Random Forest splitting rule minimizes the prediction error. Whereas, GRF's splitting rule maximizes the heterogeneity in estimated treatment effects in the terminal nodes of all trees. GRF uses a data-driven weighting function to estimate the treatment effect. Each observation is weighted based on the frequency that it falls into the same leaf as the target observation over all trees.[5] GRF solves a local Generalized Method of Moments (GMM) model using these weights by minimizing the weighted moment conditions and estimates the heterogeneous treatment effect. GRF is powerful in estimating the causal treatment effect even when the treatment variable is endogenous.

Additionally, for transporting results to an untreated population I use a weighting-based method. In this approach the population that I estimated the causal effect (e.g. Oregon experiment) is called study population and the population that I transport results to is called target population. Individuals in these two population are different based on the distributions of their characteristics. I use the proposed weighting method (inverse of

---

[5] The weighting function uses a measure of similarity. This measure is such that the closer observations to the target individual in terms of characteristics receive larger weights.

odds) by Cole and Stuart (2010), to weight individuals in the study population to resemble individuals in the target population. The propensity score used in the weighting method is defined as the probability that each individual participating in the experiment or study population. I use several methods to estimate the propensity scores, namely a linear method with Logistic Regression (LR), and more flexible non-linear Machine Learning methods including Random Forest (RF), and Gradient Boosting Method (GBM). I compare the results obtained from all these methods on a synthetic set of covariates which I generate using a Copula strategy.

1.4 Related Works

Prior studies on Oregon Health Insurance Experiment (OHIE) used the random assignment of the lottery and studied the average causal impact of Medicaid enrollment on several outcome variables. Finkelstein et al. (2012), Baicker et al. (2013), and Baicker et al. (2014) provide comprehensive analysis of the OHIE. They found that after two years that people had Medicaid insurance, the Medicaid decreased out-of-pocket medical costs, increased the healthcare utilization, improved the self-reported general health, reduced depression. Moreover, they examined the effect of Medicaid on people's physical health but did not found statistically significant effects specifically for blood pressure and cholesterol. In their analysis, the effects of Medicaid on labor market, employment, and earning were inconclusive.

The common approach between all previous studies is that they used the lottery as the instrumental variable and using the two-stage least squares method they estimated the Local Average Treatment Effect (LATE).

A few papers estimated the heterogeneous effects of Medicaid in OHIE. Finkelstein et al. (2012) in the online appendix, examine the potential heterogeneity in effects of Medicaid on hospital utilization. They found that older people have a greater increase in healthcare utilization compared to younger people. Finkelstein et al. (2012) and Taubman et al. (2014) use subgroup analysis to estimate the heterogeneous effect which is very different from the approach that I use in this study. In Chapter 2 I'll explain the problems with subgroup analysis approach. There are other recent studies on the heterogeneity of Medicaid's effect on OHIE which are in line with my approach and my findings that I explain in the following.

Qiu et al. (2021) developed a Lasso estimation under a non-convex objective function for a two-stage regression model. They studied the heterogeneity of Medicaid's effect with respect to limited number of covariates, i.e., age, race, and gender. They find that Medicaid increases happiness for females between 35 to 49 years of age and people above 50 years old irrespective of their gender. Despite the difference in the method, their results on happiness align well with my findings in this paper.

Shakya and McCullough (2020) studied the heterogeneous effect of "Access to Medicaid" on various health outcomes using the cluster-robust GRF. They restricted their attention to the heterogeneity only in two dimensions, age and income; and find that the impact of access to Medicaid is greater among older adults and poorer households. In contrast, I study the heterogeneous effect of "Medicaid Coverage" on various self-reported happiness and consider heterogeneity in multiple dimensions to shed light on underlying mechanisms through which Medicaid may affect this outcome.

1.5 Dissertation Layout

The work presented in this dissertation is organizes as follows.

In Chapter 2, I describe the Oregon Health Insurance Experiment setting and the dataset that I use for this study. Moreover, I discuss the methodology that I use. I start with this fact that individuals with and without Medicaid coverage are inherently different and we are facing with the selection bias problem. As such, the Instrumental Variable (IV) is an ideal identification strategy for estimating the causal effect of Medicaid enrollment. I describe the IV method in both the traditional framework and the potential outcome framework and explain why the traditional framework is not suited for the purpose of this study. Then I show the concerns with the traditional method of econometrics in estimating the heterogeneous effects. In the end, I introduce the Generalized Random Forest (GRF) method as a non-parametric causal Machine Learning method that estimates the causal effect of Medicaid enrollment as a function of individuals' characteristics.

In Chapter 3, I use the GRF method to estimate the heterogeneous effects of Oregon's 2008 Medicaid expansion program on four outcome variables: self-reported happiness, self-reported out-of-pocket medical costs, and self-reported depression. I show the effect heterogeneity with respect to specific characteristics. In the end, I discuss the potential underlying mechanisms deriving the estimated heterogeneous effects.

In Chapter 4, I use the estimated individual-level effects from Chapter 3 to transport them to a population where the Medicaid expansion program has not implemented yet. I use two methods for transporting results: 1) weighting-based approaches, 2) predicting effects using GRF trained on Oregon data.

Chapter 5 concludes this dissertation. A summary of this dissertation is presented, and results from the analysis are discussed.

Chapter 2

GENERALIZED RANDOM FOREST: A FLEXIBLE APPROACH TO ESTIMATE THE HETEROGENEOUS EFFECTS OF MEDICAID EXPANSION IN OHIE

2.1 Introduction

Oregon offers Medicaid coverage through Oregon Health Plan (OHP). It consists of two main programs, OHP Plus and OHP Standard and provides benefits[1] for a limited number of low-income uninsured adults, with low monthly premiums (between $0 and $20, based on income) and no cost sharing. OHP Plus provides health benefits for specific low-income groups such as senior adults older than 65, children under 19, pregnant women, blind and disabled people, and those who qualify for the Temporary Assistance to Needy Families (TANF) program.

The subject of this study is the OHP Standard program which provides Medicaid to those who are not categorically eligible for OHP Plus. This program is offered only when the State provides additional funding. The limited number of available slots are offered through a lottery offering.

In early 2002, about 110,000 individuals were enrolled in OHP Standard. Since this plan provided an expensive comprehensive benefit plan (about $3,000 per person), the State faced with budget shortfall and didn't accept any new enrollment in 2004. By early 2008, the State had enough budget to enroll an additional 10,000 adults. Because the government could not afford to enroll all those who wanted and were eligible for enrolling into the program, the State conducted a series of lottery draws from a waiting list. Through this lottery, there is a unique opportunity to study the causal impact of Medicaid for the uninsured on several outcomes with a randomized controlled design (Taubman et al., 2014).

The best identification strategy in this setting is using the Instrumental Variable (IV).

[1] It covers prescription drugs, mental health and chemical dependency services (including outpatient services), physician services, all major hospital benefits, and hospice care.

In the following sections, I describe the randomized control trial setting of the Oregon experiment and the different approaches that instrumental variable can be used to estimate the causal effects of Medicaid coverage.

2.2 Experiment Settings

In 2008, from January 28 to February 29, about 90,000 individuals signed up for the lottery list.[2] To be eligible for OHP Standard, individuals had to have income below the Federal Poverty Level[3] (FPL) and have assets below $2,000. They had to be between 19 to 64 years of age, Oregon residents, US citizens or legal immigrants, with no health insurance for at least 6 months, and not otherwise be eligible for Medicaid or other public insurance. Before conducting lottery draws, they excluded those who didn't meet all the criteria from the waiting list (Finkelstein et al., 2012; Baicker et al., 2013).[4]

After excluding the ineligible applicants, the State conducted a series of lottery draws among 75,000 eligible applicants in the waiting list. About 30,000 individuals won the lottery and were given the chance to apply for Medicaid and receive health insurance benefits for themselves and their family members. Those individuals who did not win the lottery (about 45,000 individuals) were considered as a control group. In the end, about 30% of selected individuals met all the requirements and successfully enrolled in the OHP Standard program.

2.3 Data

The participants in the study were surveyed several time at different intervals to collect the data of interest. I select my dataset from the results of these surveys. The survey results are publicly available and can be accessed through the NBER website.[5] All datasets contain

---

[2] There was no specific requirement to sign up for the lottery list.

[3] The Federal Poverty Level was $10,400 for an individual and $21,200 for a family of four in 2008.

[4] Those who gave an address outside of Oregon (36 individuals), were aged less than 19 or more than 64 years on Jan 1, 2008 (3,258 individuals), gave group or institutional address (5,161 individuals), signed up by unrelated third party (5,708 individuals), died prior to the notification date (134 individuals),and multiple active observations (605 individuals).

[5] https://www.nber.org/

observations at the individual level.

The "Descriptive Variables" survey includes demographic information that individuals reported about themselves at the time of lottery sign up (January and February 2008). Basic demographic information in this survey including gender, birth year, whether English was the preferred language of communication, whether the applicant signed up by themselves, zip code, and number of people in household were provided by individuals on the lottery sign-up list. The lottery results for each individual is also included in this dataset.

The "Initial Mail" survey was conducted shortly after the random assignment, between June 2008 and November 2008. This survey was sent to 58,405 individuals (29,589 winners and 28,816 non-winner of the lottery) and contained questions on health insurance, health care needs and health care use, and additional demographic information. Most questions asked individuals to consider the past six months (e.g., "Have you skipped bills to pay health care bills in the last 6 months?"). I use the information in "Initial Mail" survey and "Descriptive Variables" survey as baseline covariates in my study. To be consistent with the literature I selected the same covariates as Finkelstein et al. (2012) used for their analysis. These covariates include age, race, gender, education, weekly work hours, income category (as a percentage of FPL), registration on the first day, zip code, P.O. Box on record, household size, language, phone on record, and self registration. The mean of these variables for both winner and non-winner groups are provided in Table 2.1.

The "Twelve-Month Mail" survey was conducted on the same sample (58,405 individuals) one year after the initial survey and same questions were asked. I select my outcome variables, namely *Happiness*, *Out-of-Pocket Medical Costs*, *Depression* from this survey.

The data on the binary variable *Happiness* is acquired through the answer to the question "What is your overall happiness?". It takes a value of 0 if the respondent indicates that they are not too happy, and takes a value of 1 if the respondent indicated that they are very happy, or pretty happy. The *Out-of-Pocket medical costs* takes values 0 or 1, with respect to the answer to "Any out of pocket cost for medical care in the past 6 months?". The positive responses are indicated by the value of 1 and negative responses are indicated by 0. The *Depression* variable is the answer to the question "How often have you felt down, depressed, hopeless in the past 2 weeks?". This variable is measured from 1 (not at all) to 4

Table (2.1) **Summary Statistics**

This table shows the summary statistics of variables that I use in the model. Variables with their different levels are shown in the first and second columns, respectively. The third and forth columns show the mean of the variables for lottery winner and non-winner populations. The last column shows the *p*-value of the t-test for difference in mean of each variable among lottery winner and lottery non-winner people.

Variable	Level	Winner Mean	Non-Winner Mean	*p*-value
Age	19-30	18.4%	18.5%	0.81
	30-40	20.4%	18.2%	0.00*
	40-55	40.2%	41.4%	0.10
	55-64	20.8%	21.6%	0.20
Race	White	74.4%	75.0%	0.32
	Black	2.3%	2.4%	0.82
	Hispanic	1.9%	13.7%	0.00*
	Others	21.4%	8.9%	0.00*
Gender	Female	59.8%	61.5%	0.02
	Male	40.1%	38.4%	0.02
Education	Less than high school	17.0%	16.8%	0.68
	High school disploma or GED	49.9%	50.6%	0.41
	Vocational training or 2-year degree	21.6%	21.2%	0.54
	4-year college degree or more	11.3%	11.2%	0.97
Employment	Do not currently work	52.0%	52.7%	0.38
	Work <20 hours per week	0.09%	0.09%	0.62
	Work 20–29 hours per week	10.8%	11.1%	0.50
	Work 30+ hrs per week	27.8%	26.5%	0.07
Income (% FPL)	<50%	35.4%	37.5%	0.00*
	50–75%	11.3%	10.7%	0.23
	75–100%	13.4%	13.2%	0.77
	100–150%	15.9%	14.7%	0.02
	Above 150%	10.8%	9.7%	0.02
ZIP code	MSA	74.2%	74.3%	0.88
	Not MSA	25.7%	25.6%	0.88
Language	English preferred	92.1%	92.9%	0.03
	Not Englisg preferred	7.8%	7.0%	0.03
Registration on the first day	Yes	10.9%	10.7%	0.62
	No	89.0%	89.2 %	0.62
Self registration	Yes	84.2%	88.5%	0.00*
	No	15.7%	11.4 %	0.00*
Provide phone number	Yes	90.6%	89.9%	0.10
	No	9.3%	10.1 %	0.10

Note: *p<0.1; **p<0.05; ***p<0.01

(nearly every day). Following Finkelstein et al. (2012) I select the treatment variable, *Ever enrolled in Medicaid during the study period*, from the "State Program Variables" survey.

Of the 58,405 participants in the surveys, 26,423 and 23,741 individuals responded to the "Initial Mail" and "Twelve-Month Mail" surveys, respectively. As the response rate is less than 50%, the information about the respondents may not be representative of the full surveyed sample (i.e., non-response bias). In other words, it is highly probable that more treated (lottery winner) or control (lottery non-winner) people select to respond to surveys and this differential selection ruins the randomized structure of the experiment. In order to address this concern, in Table 2.1, I show that means of the characteristics for both groups of lottery winners and lottery non-winners are very similar. Moreover, the p-value of the t-test for the mean difference of each variable is reported in the last column of Table 2.1. Except for six levels of variables, the reported p-values show that the lottery assignment is random at 99% confident level. Therefore, I can rely on the randomized property of the experiment and take advantage of the created balanced samples to study the impacts of Medicaid.

In all surveys, individuals are identified by a "Person ID". The Person ID is used to merge and consolidate information for each person from different surveys as I combined the surveys. After cleaning the data I found that about 16,000 unique Person IDs are in common among all surveys.

2.4 Empirical Methodology

Ideally, we would randomly assign Medicaid to individuals and study its effect by comparing the outcome of the treated and the control groups. But in reality we have self-selection bias in Medicaid enrollment. The probability of enrolling in Medicaid program is correlated to many factors including: institutional factors such as law, and requirements for Medicaid eligibility; individual factors such as age, education, health status, and culture; and the provider factors such as medical costs, or the quality of services. Any of these factors can encourage or discourage individuals from enrolling in Medicaid programs and rule out the random assignment of Medicaid. For example healthy people may be less likely to enroll in the Medicaid program. This may result in a negative correlation between Medicaid enrollment and health quality.

In a regression model, in order to estimate the effect of Medicaid on outcome variable and to address the selection bias problem we have to control for all the mentioned factors. However, some factors such as culture, quality of services, individuals' attitude, etc. are not observable while they may affect both Medicaid enrollment and the outcome variables simultaneously. This introduces endogeneity in the variables of our interest. Random assignment of Medicaid subscription would allow us to easily control for these unobservable factors. With random assignment of Medicaid, subscribers and non-subscribers are distributed similarly in all dimensions except in Medicaid subscription. OHIE is a randomized controlled experiment where the lottery drawings for access to Medicaid can be considered as an exogenous source of randomness to evaluate the effect of Medicaid on health outcomes.

Angrist and Imbens (1995) proposed Instrumental variable (IV) method which can handle endogeneity and deliver unbiased estimate of the Medicaid effect. The next section explains the IV concept.

2.4.1 Instrumental Variable

The Oregon experiment design is a rare opportunity to address the endogeneity of Medicaid enrollment by using an Instrumental Variable (IV) which is correlated to Medicaid, but uncorrelated to the outcome. The Instrumental Variable is an exogenous source of randomness driving the treatment, without affecting the outcome.

The idea behind the IV identification is as follows. In observational studies, the unobserved confounding factor affects both the treatment variable and the outcome variable at the same time. As such, we cannot estimate the effect of the treatment on outcome variable directly as this effect is contaminated by the influence of a confounder. To solve this problem, we need to find an exogenous variable such that its variation affects the treatment but it has no effect on the outcome directly. This variable is the instrumental variable and its variation is assumed to be random. Therefore, using IV, the estimated effect can be considered solely due to treatment and be interpreted as the causal effect of treatment on the outcome.

In general, there are three main assumptions for IV identification: (1) Exogeneity of the IV, (2) Exclusion restriction, i.e., IV does not directly affect the outcome variable; it only affects the outcome through the treatment variable, (3) Relevance, i.e., variation in IV causes variation in the treatment variable.

The OHIE provides a unique situation in which an exogenous variable is readily available for evaluating the causal effects of Medicaid enrollment on various outcome variables. The lottery draw that is used for access to Medicaid subscription in the OHIE can be used as an IV. That is the case because it is random (it satisfies Assumption 1), it does not have direct effect on outcome variables (i.e., winning the lottery is completely random and unrelated to the outcome, so it satisfies Assumption 2) but it has direct effect on the treatment variable (that is if someone wins the lottery, he or she can sign up for Medicaid, also satisfying Assumption 3).

In what follows, I explain two approaches for using IV in treatment effect estimation when treatment variable is endogenous. Each of these approaches has an additional assumption besides the ones that I mentioned above.

2.4.2 Instrumental Variable in Regression Models

Instrumental Variable method uses the structural model in the traditional econometrics framework.[6] In the traditional model, the relationship between the endogenous treatment variable, the exogenous covariates, and the outcome variable is linear. The IV identification is based on the Two-Stage Least Squares (2SLS) estimator. The two stages of IV estimation in my case are

$$\text{Medicaid}_i = \delta_0 + \delta_1 \text{Lottery}_i + \delta_2 X_i + \nu_i \tag{2.1}$$

$$\text{Outcome}_i = \alpha_0 + \alpha_1 \widehat{\text{Medicaid}}_i + \alpha_2 X_i + \varepsilon_i \tag{2.2}$$

[6] The IV, for the first time, was introduced by Philip G. Wright and possibly his son Sewall Wright in his 1928 book "The Tariff on Animal and Vegetable Oils"(Epstein (1989), and Stock and Trebbi (2003)).

where Medicaid_i is an indicator variable for treatment. It takes a value equal to 1 if an individual wins the lottery, completes the application form, meets all the requirements, and receives Medicaid coverage, whereas it takes a value equal to 0 if they do not receive Medicaid coverage. Lottery_i is an indicator variable which denotes whether the individual i was selected by the lottery. X_i is a vector of observed covariates that are correlated with treatment probability and outcome. Equation (2.1) denotes the first stage where the endogenous variable is regressed on all of the exogenous variables including the instrument, i.e., Lottery_i. In the second stage, equation (2.2), the outcome is regressed on predicted value of Medicaid from the first stage and other covariates. In this context, α_1 is the variable of interest which estimates the average causal effect of Medicaid on various health outcomes. The additional assumption in this framework is that the treatment effect is homogeneous across individuals and the estimated α_1 represents the average effect for a marginal individuals in the population. Almost all papers on OHIE for evaluating the Medicaid expansion use the 2SLS method and report the average effect of Medicaid on various outcome variable (e.g., Finkelstein et al., 2012; Baicker et al., 2013; Taubman et al., 2014; Finkelstein et al., 2016; Allen et al., 2013; Baicker et al., 2017, 2014). Likewise, I estimate the effect of Medicaid on selected outcome variables using the 2SLS method and report the average effects for each outcome variable in Table 2.2.

Table (2.2) **Average Effect of Medicaid on Selected Outcome Variables**
In this table I run the 2SLS regressions using the covariates in Table 2.1 for three selected outcome variables. For each regression I only report the coefficient of Medicaid, its standard error in parenthesis, and the corresponding p-values.

Outcome	Medicaid's Effect	p-values
Happiness	0.20***	
	(0.02)	0.0
Out-of-Pocket Medical Costs	-0.22***	
	(0.02)	0.0
Depression	-0.21***	
	(0.05)	0.0
Note:	*p<0.1; **p<0.05; ***p<0.01	

The homogeneity assumption implies that each individual in the sample is equally impacted by Medicaid. But homogeneity seems an unrealistic assumption in most settings and estimating the average effect in a population where individuals respond in different ways to a change in treatment, is not helpful. To relax this assumption and allow for heterogeneity, I discuss IV in potential outcome framework in the next section.

2.4.3 Instrumental Variable in Potential Outcome Framework

The most popular approach to estimate causal effect is based on the potential outcome framework (Sekhon (2008), Holland (1986)). Consider a setup where an N-unit sample, indexed by $i = 1, \cdots, N$ is drawn from a population with equal probability. Each unit has a vector of features X_i, and a real-valued scalar outcome $Y_i \in \mathbb{R}$. In Neyman-Rubin potential outcome framework, we focus on binary endogenous treatment and binary IV, which are denoted by $D \in \{0, 1\}$ and $Z \in \{0, 1\}$, respectively. In this setting, D_i depends on Z_i and I have two potential treatments.

$$D_i(j) = D_i(Z_i = j) \text{ for } j = 0, 1$$

In this study, $Z_i = 1$ if an individual wins the lottery, and $D_i(1) = 1$ if a person wins the lottery and decides to complete the application process and eventually enrolls in Medicaid. As such, the observed treatment can be written as

$$D_i = Z_i D_i(1) + (1 - Z_i) D_i(0) \tag{2.3}$$

The potential outcome for each individual is noted by $Y_i(d, z)$, where individual i had received treatment $D_i = d$ and encouragement[7] $Z_i = z$. The relationship between realized and potential outcomes can be shown by equation (2.4).

$$Y_i = D_i Y_i(1, z) + (1 - D_i) Y_i(0, z) \tag{2.4}$$

Given the exclusion restriction assumption, the potential outcomes are independent of

---

[7] Here we use "treatment assignment", "lottery", and "encouragement" interchangeably.

the instrument, i.e., $Y_i(1, z) = Y_i(1)$ and $Y_i(0, z) = Y_i(0)$.

As mentioned previously, the homogeneity assumption in traditional IV framework is not a plausible assumption in a population where individuals with different characteristics are affected differently by the treatment. Consequently, individualized treatment effect, i.e., $\tau_i = Y_i(1) - Y_i(0)$, is ideal for policy making purposes in heterogeneous population.

However, τ_i cannot be identified, because only one of the potential outcomes, $Y_i(1)$ or $Y_i(0)$, is observed for each individual at a time. As such, I cannot estimate the treatment effect for each individual by comparing the potential outcomes in the two states, i.e., treated and untreated states. This is why researchers generally resort to Average Treatment Effect (ATE) although ATE is not the ideal parameter.

Since we are interested in the heterogeneity of the treatment effect, ATE is not our parameter of interest. Instead, we estimate the Conditional Average Treatment Effect (CATE) which is a parameter that can capture heterogeneity. The CATE is interpreted as the average treatment effect for a subgroup of individuals with characteristics, $X_i = x$. The formal representation of the CATE is shown in equation 2.5.

$$\tau(x) = E[Y_i(1) - Y_i(0)|X_i = x] \tag{2.5}$$

$\tau(x)$ is a function of x and is a measure of heterogeneous effect. As we allow for heterogeneity in this framework, it is reasonable to assume that people have different reactions to the treatment assignment (IV). Angrist and Imbens (1995) suggested to divide the population into four subgroups based on the treatment assignment and the treatment take-up in the case where both the treatment and the IV are binary. The first group is "Always-takers", who take the treatment regardless of the encouragement. Conversely, the second groups is the "Never-takers", who never take the treatment even if they are assigned to the treatment group. The third group is called the "Compliers", who take the treatment only if they are assigned to the treated group and do not take the treatment otherwise. The fourth groups is the "Defiers", who behave exactly opposite to the compliers. Figure 2.1 shows these fours subgroups and their treatment status.

In addition to the three main assumptions for IV, we need to make the additional as-

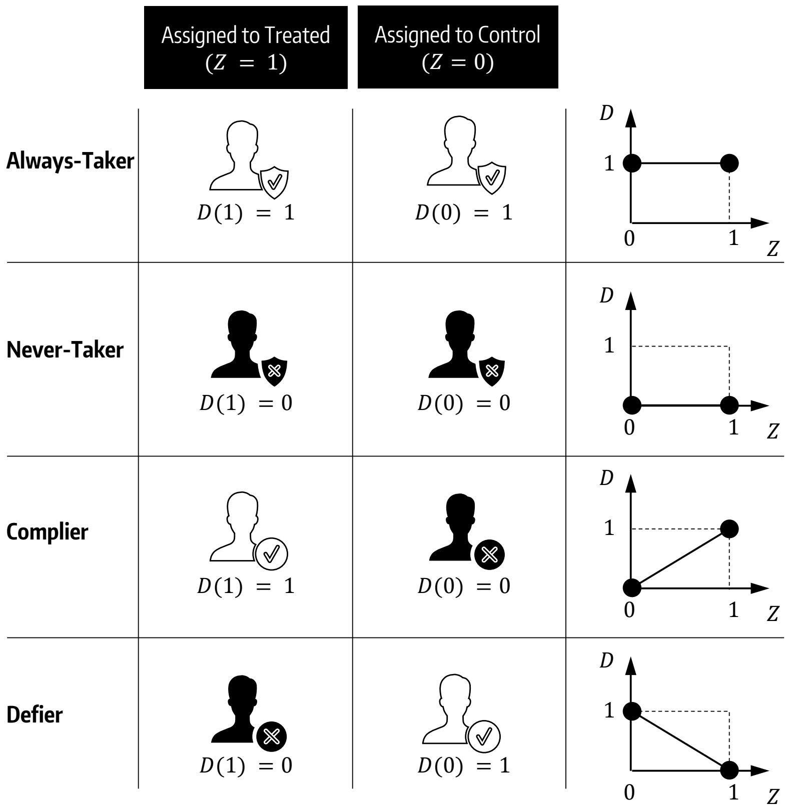

Figure (2.1) **Subgroups Based on Treatment Assignment and Take-up**
The population is divided into four subgroups based on the treatment assignment and how the individuals react to it. Always-takers take the treatment in all situations; never-takers reject the treatment regardless of assignment; compliers take the treatment only if they are assigned to the treatment; and defiers behave opposite to the compliers. Monotonicity assumption rules out the existence of defiers.

sumption of monotonicity in order to estimate the conditional average treatment effect. Angrist et al. (1996) introduced the monotonicity assumption for IV identification which states that the presence of the instrument never dissuades someone from taking the treatment. As such, defiers are excluded from the population by definition. We can write the monotonicity assumption in terms of the treatment assignment and treatment take-up as in equation 2.6.

$$D_i(1) - D_i(0) \geq 0 \tag{2.6}$$

Given the four assumptions for IV in potential outcome frame work (exogenous IV, exclusion restriction, relevance, and monotonicity), if there is a shift in IV, in the absence of defiers, the compliers are the only group whose treatment choice is affected by the IV. This is illustrated in Figure 2.1 on the right panel where D is plotted with respect to Z for each subgroup. Therefore, when we estimate the effect of the treatment variable using IV, we are estimating the effect only for the compliers. As such, this effect is called Complier Average Effect or Local Average Treatment Effect, LATE (Angrist and Imbens (1995)).

Angrist and Imbens (1995) show that under the four assumptions for IV identification within the potential outcome framework, the Wald estimator estimates causal treatment effect. But the estimated causal effect represents the effect specifically for the compliers. Therefore, the IV estimation in this setting recovers the Local Average Treatment Effect, i.e., the causal effect for the compliers. The LATE theorem[8] states that

$$\text{LATE} = E[Y_i(1) - Y_i(0)|D_i(1) \geq D_i(0)] = \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]} \qquad (2.7)$$

In the case of our study, Equation 2.7 is the causal effect of Medicaid enrollment for compliers. That is, the causal effect of Medicaid on those who sign up for Medicaid only if they win the lottery. Therefore, if we assume the treatment effect is heterogeneous across individuals, then ATE and LATE estimate two different parameters.

Even though LATE is the estimated effect for the complier group only, we cannot identify complier individuals, because for each individual we can only observe one of the potential treatment take-up ($D_i(1)$ or $D_i(0)$). For example, when $D_i = 1$ and $Z_i = 1$, compliers and always-takers are not distinguishable. Similarly when $D_i = 0$ and $Z_i = 0$, compliers and never-takers are not distinguishable.

Estimating heterogeneous treatment effect using the IV identification within the potential outcome framework leads to the conditional local average treatment effect, CLATE. In other words, calculating LATE exclusively for individuals with similar covariates resulting in conditional local average treatment effect or CLATE. Equation 2.8 presents this.

[8] The proof can be found in Angrist and Imbens (1995).

$$\text{CLATE} = \text{LATE}(x) = E[Y_i(1) - Y_i(0) | D_i(1) \geq D_i(0), X_i = x] \tag{2.8}$$

To summarize, in the traditional IV framework, the effect of treatment is assumed constant for all compliers and non-compliers. Therefore, LATE and ATE are equal. But in the potential outcome framework, IV identification recovers the Local Average Treatment Effect, LATE, or the Average Treatment Effect for the compliers within the population. Conditional Average Treatment Effect, CATE, is a parameter that can capture the heterogeneity within the population with regards to individual characteristics. Using IV identification to calculate the Conditional Average Treatment Effect leads to Conditional Local Average Treatment Effect, CLATE, which captures the heterogeneous effect for the compliers in the population.

In the following section I describe the traditional approach for estimating heterogeneous treatment effect in econometrics and identify its limitations which eventually leads us to the Generalized Random Forest (GRF) method using which we can calculate CLATE for several outcome variables.

2.5 Commonly Used Methods in Estimating Heterogeneous Treatment Effect

In econometrics, the typical way of finding whether a treatment effect is heterogeneous with respect to an observable characteristic, e.g., X_1, is by including the interaction term of the treatment indicator and X_1 in the regression model. For instance, assume that I am interested in estimating the effect of Medicaid on hospital utilization and I want to test whether this effect is heterogeneous among individuals with different age ranges. To do so, I include the interaction term of Medicaid indicator and the age variable. If the coefficient of the interaction term is statistically significant, I conclude that the effect of Medicaid coverage on hospital utilization depends on age. If I go further and want to test whether the effect of Medicaid varies with respect to age and race, I need to include the interaction terms of Medicaid indicator and each variable (age and race) and also the interaction term of age and race and Medicaid indicator as shown in the equation below.

$$\begin{aligned}\text{Hospital Utilization} = \alpha_0 &+ \alpha_1\text{Medicaid} + \alpha_2\text{Age} + \alpha_3\text{Race} + \alpha_4\text{Age} \times \text{Race} \\ &+ \alpha_5\text{Medicaid} \times \text{Age} + \alpha_6\text{Medicaid} \times \text{Race} \\ &+ \alpha_7\text{Medicaid} \times \text{Age} \times \text{Race} \qquad (2.9)\end{aligned}$$

In this example, with 2 covariates, I need to estimate 8 parameters. The number of parameters grow exponentially with the number of covariates used in the model. That is, if I want to use 10 variables I have to estimate 2048 parameters. This method quickly becomes problematic when there are many covaraites, because, for detecting heterogeneity in an accurate way, I have to include the interaction of all possible combinations of covariates with and without the Medicaid indicator. In the general case with p covariates I need to estimate 2^{p+1} parameters. As such, using the traditional method we can find heterogeneity in very few dimensions. In addition to that, the correct functional form of the regression is unknown, i.e., whether I have to include Age $\times$ Age, or Age $\times$ Age $\times$ Age, etc., to the model or not.

To address these problems, Athey et al. (2019) proposed a machine learning method which non-parametrically considers all interaction terms and estimates the heterogeneous causal effect of an endogenous variable. In the next section I discuss the GRF method.

2.6 Generalized Random Forest

The individual treatment effect, $\tau_i = Y_i(1) - Y_i(0)$, can never be observed since only one of the potential outcomes is realized for each individual. Therefore, it is impossible to estimate τ_i using standard supervised machine learning algorithm.

Generalized Random Forest (GRF) developed by Athey et al. (2019) is a non-parametric statistical estimation method, which is based on the Causal Random Forest (CRF) method (Athey and Imbens, 2016; Wager and Athey, 2018). GRF extends the idea of CRF by eliminating the restriction on the randomized treatment assignment. CRF builds on the idea of Causal Tree (CT). The causal tree algorithm recursively splits the covariate space into partitions. Then within each partition it estimates the treatment effect. This treatment effect is calculated using the following equation.

$$\hat{\tau}(x) = \frac{\sum\limits_{X_i \in l(x), D_i=1} Y_i}{\sum 1[X_i \in l(x), D_i = 1]} - \frac{\sum\limits_{X_i \in l(x), D_i=0} Y_i}{\sum 1[X_i \in l(x), D_i = 0]} \tag{2.10}$$

In this equation $l(x)$ denotes a leaf of the tree containing a set of observations with covaraites that are very similar to the target point, x. The partitioning is performed to capture all the heterogeneity in treatment effect.

GRF modifies the random forest method, and estimates any heterogeneous quantity, $\theta(x)$, from the data by performing a weighted kernel estimation. The closest observations to the target observation receive a larger weight and those less similar to the target observation, receive smaller weights. GRF estimates $\theta(x)$ using local moment condition:

$$E[\Psi_{\theta(x),\nu(x)}(O_i)|X_i = x] = 0, \quad \forall x \in \chi \text{ and } i = 1, \cdots, N \tag{2.11}$$

Where $\Psi(\cdot)$ is a scoring function, $\nu(x)$ is an optional nuisance parameter, X_i is a vector of covaraites, and O_i is the set of observables. In my study, $O_i = \{Y_i, D_i, Z_i\}$, where Y_i refers to various outcomes, D_i is a binary variable which denotes whether an individual i has Medicaid coverage, and Z_i is an indicator variable that represents whether someone won the lottery or not. The parameter of interest is $\theta(x)$ which represents the estimated causal effect of Medicaid on various outcomes. $\theta(X_i)$ is a function of the covaraites and captures the heterogeneous treatment effect. The structural model in Equation 2.12 denotes the relationship between outcome Y_i, endogeneous treatment D_i, and the noise term, ε_i, that is correlated with the treatment variable:

$$Y_i = \nu(X_i) + \theta(X_i)D_i + \varepsilon_i \tag{2.12}$$

It is straightforward to show that $\theta(X_i)$ represents the CLATE. Since in Equation 2.12, ε_i is correlated with the treatment variable I need to use the IV, Z_i, as an exogenous variable to deal with the endogeneity in Medicaid. In what follows I show how GRF addresses heterogeneity and then I explain IV identification using GRF method.

In a random forest, the criterion for splitting the covariate space for predicting the for the target observation, is minimizing the difference between the true and the predicted value. But in case of estimating the causal effect of a variable, the true value of the parameter is not observed, so the error term cannot be minimized. In contrast, split selection algorithm in GRF emphasizes maximizing heterogeneity in treatment effect.

In GRF there is not a loss function to be minimize. Instead, it solves moment conditions. The Equation 2.11 is the general form of the moment condition that is solved for $\theta(x)$ and $\nu(x)$ such that the moment condition get as close as possible to zero. In the context of IV identification, the moment condition used in the GRF algorithm comes from the exclusion restriction condition, i.e., $E[Z_i \cdot \varepsilon_i | X_i = x] = 0$, and the orthogonality of errors condition, i.e., $E[\varepsilon_i | X_i = x] = 0$. Therefore, the Ψ function is set as follows:

$$\Psi_{\theta(x),\nu(x)} = \begin{pmatrix} Z_i \left(Y_i - \theta(x)D_i - \nu(x)\right) \\ Y_i - \theta(x)D_i - \nu(x) \end{pmatrix} \tag{2.13}$$

To derive parameters $(\theta(x), \nu(x))$ the following weighted moment conditions need to be minimized:

$$(\hat{\theta}(x), \hat{\nu}(x)) \in \underset{\theta,\nu}{\arg\min} \left\{ \left\| \sum_{i=1}^{N} \alpha_i(x) \Psi_{\theta,\nu}(O_i) \right\|_2 \right\} \tag{2.14}$$

where $\alpha_i(x)$ is the weighting function which is derived from a Random Forest. The core idea behind GRF is that in estimating parameters for a particular value of covariates, nearby observations in the covariate space get higher weights. These weights are calculated from the frequency with which the i^{th} training observation falls into the same leaf as the target value of the covariate vector. Formally, $\alpha_i(x)$ is given by:

$$\alpha_i(x) = \frac{1}{B} \sum_{b=1}^{B} \alpha_{bi}(x) \quad \text{where} \quad \alpha_{bi}(x) = \frac{\mathbf{1}(X_i \in l_b(x))}{|l_b(x)|} \tag{2.15}$$

where B is the number of trees and $l_b(x)$ is the partition that x belongs to, in the b^{th} tree.

After finding the data-driven weights using Random Forest and solving for the weighted moment condition in Equation 2.14, the estimated $\theta(x)$ represents the Conditional Local Average Treatment Effect (CLATE).

2.7 GRF Advantages to Traditional Econometrics Methods

GRF has several advantages over the other methods for estimating the heterogeneous treatment effect. First, GRF estimates multidimensional heterogeneities. Whereas in the method of including the interaction term between treatment variable and other covariates to the regression model, we can only estimate the heterogeneous effect with respect to few covariates. By increasing the number of covariates the number of parameters that need to be estimated easily gets out of hand. Second, GRF estimates the effect nonparametrically and utilizes flexible model selection. In the regression method with interaction terms the correct functional form of the regression model is unknown. The GRF learns the non-linearity patterns in the data and considers all the combinations of interaction terms. Third, in GRF method all the estimated heterogeneous effects come from the same model. If we use subgroup analysis (Taubman et al., 2014) for estimating the heterogeneous effects, we need to run separate regression models for distinct heterogeneities. As such, we need to compare the outcomes of different models and due to multiple hypothesis testing problem (Lan et al., 2016; List et al., 2019) the p-values of single hypothesis tests are not valid.

2.8 GRF Outputs

As I explained in previous section, the estimated effects by GRF are a function of individuals' characteristics. In other words, it provides the estimated effects at the individual-level. Since people have different characteristics, the estimated effects for each person would be different and it is called heterogeneous effects with respect to observable characteristics.

In addition to the heterogeneity in the effect, GRF provides the importance matrix of baseline characteristics. In this matrix variables are sorted based on their contribution in explaining the heterogeneity in effect. The "importance value" for each characteristics is the frequency that each characteristics was split on at each depth in the forest.

Using the estimated heterogeneous effects and the important variables I show how the effects vary with respect to the most important variables in several dimensions. In Chapter 3 I show the GRF outputs for several outcomes in Oregon experiment.

Chapter 3

MEDICAID EFFECTS: WHAT WE LEARN FROM STUDYING THE HETEROGENEITIES

3.1 Introduction

Of the several available outcome variables in the OHIE dataset, I select self-reported happiness, self-reported out-of-pocket medical costs, and self-reported depression — henceforth happiness, out-of-pocket medical costs, and depression — to estimate the heterogeneous effects of Medicaid. The underlying model in GRF, that I use to estimate each of the selected outcome variables, is as follows:

$$\text{Outcome}_i = \nu(X_i) + \theta(X_i)\text{Medicaid}_i + \varepsilon_i \tag{3.1}$$

where for the individual i, X_i is a vector of characteristics, $\theta(X_i)$ is the causal effect of Medicaid or the Conditional Local Average Treatment Effect (CLATE), $\nu(X_i)$ is a nuisance parameter, and ε_i is a noise term which is positively correlated with Medicaid. $\theta(X_i)$ captures the heterogeneity of the effect, as it is a function of the covariates.

The next few sections present the GRF results for happiness in detail. I show the heterogeneity in the effect of Medicaid on happiness in several dimensions. For the out-of-pocket medical costs and the depression variables, I will selectively demonstrate the heterogeneous effect with only a choice of few results presentation.

3.2 Self-reported Happiness

3.2.1 Distribution of Medicaid's Effect on Happiness

Happiness, as a well-being measure, is a binary variable which was acquired via the answer to the question "What is your overall happiness?" in the follow-up survey. This survey was conducted one year after the lottery drawings. The answer takes a value of 0 if the respondent indicated that they are not too happy, and a value of 1 if they indicated that they are either very happy, or pretty happy. Figure 3.1 shows the histogram of the estimated

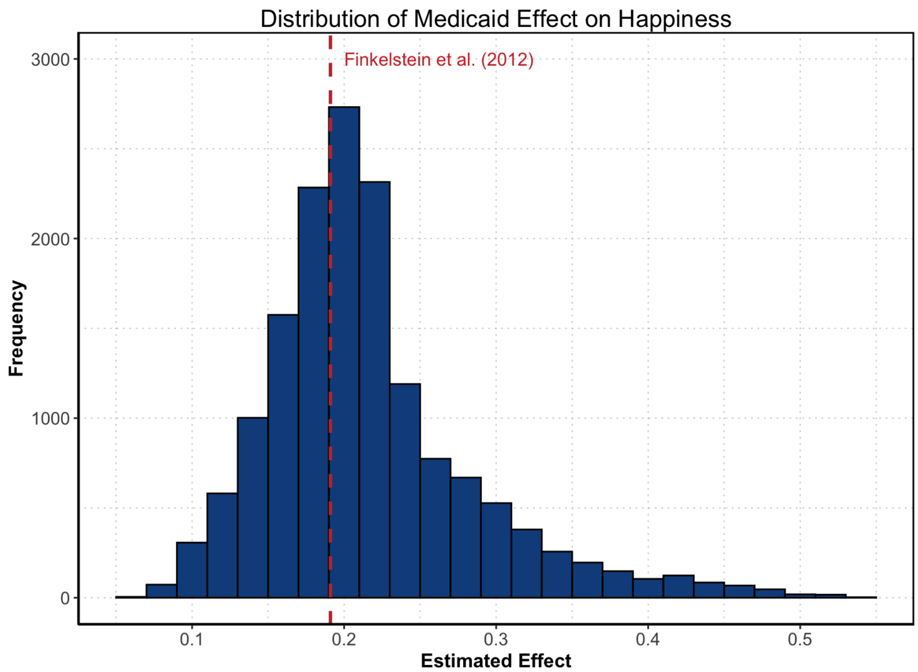

Figure (3.1) **Histogram of Estimated Medicaid's Effect on Happiness Using GRF** This plot shows the distribution of individual-level estimated effect of Medicaid on happiness. The effect for all individuals is positive and varies between 0.04 and 0.57. The mean of this distribution is 0.20. The red dashed line shows the estimated average effect of Medicaid on happiness reported in Finkelstein et al. (2012).

individual-level effect of Medicaid on happiness ($\theta(X_i)$ in Equation 3.1). As shown in Figure 3.1, those who enrolled in the Medicaid expansion program are generally happier compared to those who did not. This result is consistent with the average effect reported by Finkelstein et al. (2012). Per their report, Medicaid expansion in Oregon has increased the happiness of the program-subscribers by 0.19 after the first year. The average of the estimated individual effects from my GRF results yields a value of 0.20, which is very close to Finkelstein et al. (2012)'s result.

3.2.2 Variable Importance for Happiness

In addition to the individual-level effects, GRF provides the control variables for which the effect is most heterogeneous. For happiness, among all control variables, age, weekly working hours, and income category (as a percentage of Federal Poverty Level) are the top three most important covariates in estimating heterogeneous treatment effect.

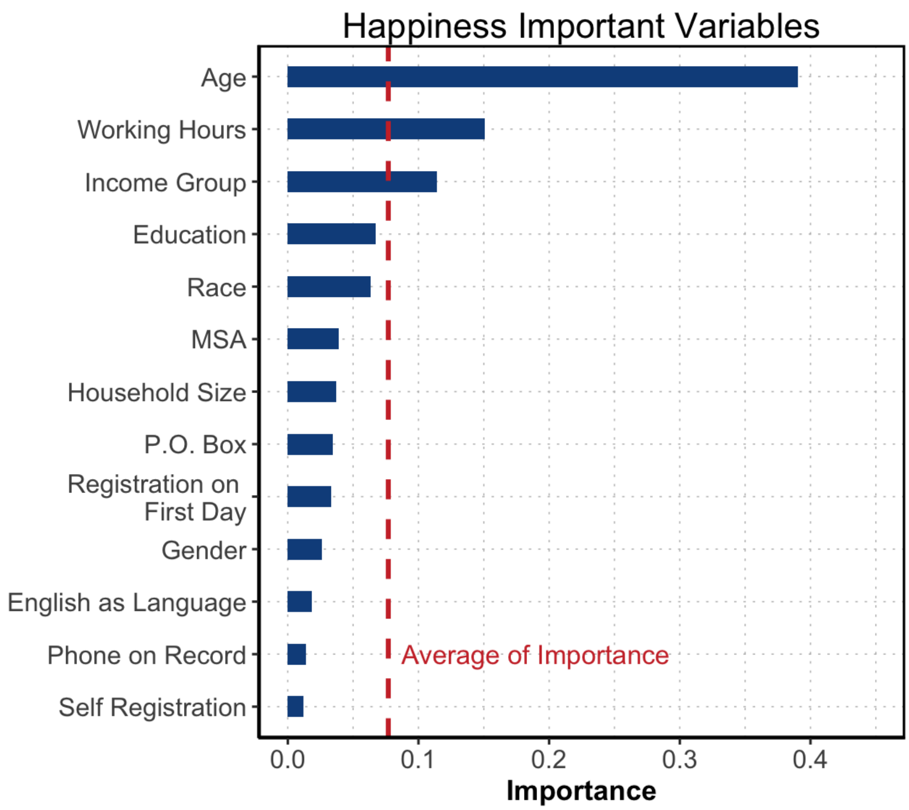

Figure (3.2) **The Importance Matrix for Happiness from the GRF Model**
This figure represents the importance measure of each variable in the GRF algorithm in estimating the effect of Medicaid on happiness. It determines how often the GRF algorithm chooses a variable for splitting the covariate space across all trees in the forest.

Figure 3.2 shows the importance values for all covariates considered in estimating the Medicaid's effect on happiness. The "importance value" is the relative frequency of each covariate being used by that GRF to split the covariate space in growing a tree. For example as shown in Figure 3.2 the importance value for the age is 0.38, for which the interpretation is that in 38% of splits, GRF selects the age variable. I use the variable importance to select the covariates that I should focus on with respect to the heterogeneity in effect. Also shown in Figure 3.2 is the average importance value with a red dashed line. Following Athey and Wager (2019) I use the average importance value as a threshold for selecting the most important variables. The top three important variables as discussed before, all have an importance values above average. In the next section, I use these most important variables to show the heterogeneity in the effect of Medicaid on happiness of the individuals.

3.2.3 Individual-Level Heterogeneity for Happiness

Heterogeneity by Age. As depicted in Figure 3.2, GRF results show that the CLATE on happiness is most heterogeneous over the age of the Medicaid participants.

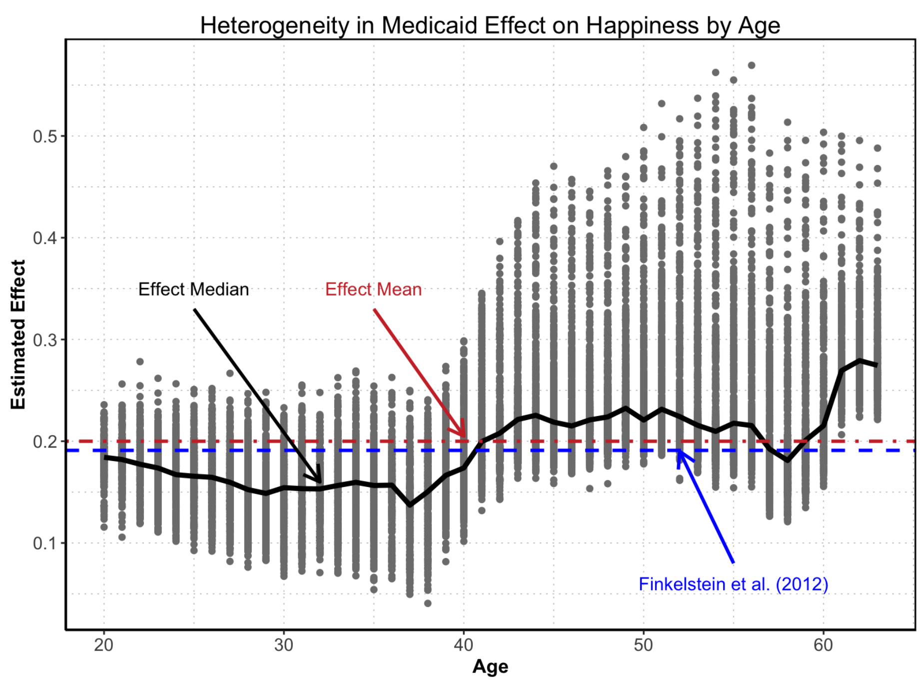

Figure (3.3) **Heterogeneity in Medicaid's effect on Happiness by Age**
This figure represents how the estimated effect of Medicaid on happiness varies with respect to age. To have a better sense of effect's variability, I added the effect's median line (the solid black line) across the age range to this figure. Additionally, this figure shows how close are the average of estimated individual effect with GRF (the red dash-dot line) and the Finkelstein et al. (2012)'s result (the blue dashed line).

To illustrate this, Figure 3.3 shows a scatter plot of the estimated CLATE versus the age of the individuals. The effect is positive for all individuals meaning that Medicaid has improved the happiness among those who are covered by Medicaid compared to those who are not. The blue dashed line in Figure 3.3 shows the average effect of Medicaid on happiness reported by Finkelstein et al. (2012). The average of the estimated individual-level effect (0.20) is also shown as a dash-dot line in red.

The median of the effect in each age level is shown in the figure with a solid black line. The effect is stronger for older people, especially those who are more than 40 years old, compared to younger people. This result confirms the findings in Qiu et al. (2021) who also concluded that the effect of Medicaid expansion in Oregon on happiness is heterogeneous across different age groups where females between 35 to 49 years old and people above 50 years old have the strongest effect compared to other age groups.[1] Conversely, the study on the impacts of Medicaid in Massachusetts' healthcare reform on self-reported happiness using subgroup analysis by Kim and Koh (2018) shows that Medicaid has improved the happiness among younger individuals more than older ones, while the effect is positive for all individuals. Moreover, Baicker et al. (2013) show that Medicaid expansion in OHIE has no significant effect on individuals' happiness. The difference between their results and mine could be attributed to using two different data sources. My outcome variable is extracted from the "Twelve-Month Mail" survey whereas their outcome variable is from the "In-person" survey.

Heterogeneity by Weekly Working Hours. The second most important variable generating heterogeneity in CLATE for happiness is the weekly working hours. This variable consists of four categories: 1) do not currently working , 2) working less than 20 hours, 3) working between 20 and 30 hours, 4) and working more than 30 hours.

Figure 3.4 shows the violin plot of estimated CLATE for each level of the weekly working hours. As shown in Figure 3.4, the distributions of effect are different for different levels of working hours, specifically for the group who work more than 30 hours per week. In order to test for heterogeneity in effect between any two-combination of the weekly work-

[1] Qiu et al. (2021) categorized the age variable into three groups: less than 35, from 35 to 49 and from 50 to 64.

ing hours categories, I use Kolmogorov-Smirnov (K.S.) two-sample test and the K-sample Anderson–Darling (A.D.) test. The null hypothesis in the K.S. test is that both groups are drawn from populations with identical distributions. The violation of the null hypothesis can be interpreted as difference in the medians, variances, or the shape of the distributions of two populations. While similar, the A.D. test is more sensitive to difference in the tails of the distribution as compared to the K.S. test.

Figure (3.4) **Heterogeneity in Medicaid's Effect on Happiness by Working Hours** In this figure the distribution of Medicaid's effect on happiness is shown with violin plot for each category of the weekly working hours. The distribution of effect for those who work more than 30 hours per week is different from other categories.

Table 3.1 shows the results of K.S. and A.D. tests. In this table, the numbers above the diagonal shows the p-values of the K.S. test and the numbers below the diagonal shows the p-values of the A.D. test for any two-combination of weekly working hours categories. Except for categories 2 and 3, the p-values for K.S. and A.D. tests show that the distribution of the effect for the weekly working hours categories are statistically different from each other at 99% confidence level. This can be visually confirmed by looking at the cumulative distri-

Table (3.1) **Kolmogorov–Smirnov and Anderson–Darling Tests**
This table shows the p-values of the Kolmogorov-Smirnov and Anderson–Darling tests. The numbers above the diagonal represents the p-values of the K.S. test and the numbers below the diagonal shows the p-values of the A.D. test for any two-combinations of weekly working hours categories. Based on p-values, I show that except for those who work less than 20 hours and those who work 20-30 hours the distribution of the effect for different weekly working hours categories are statistically different from each other

	Do Not Currently Working	Working < 20 Hours	Working 20-30 Hours	Working > 30 Hours
Not Working	–	$p_{\text{K.S.}} = 0$	$p_{\text{K.S.}} = 0$	$p_{\text{K.S.}} = 0$
Working <20 Hours	$p_{\text{A.D.}} = 0$	–	$\mathbf{p_{\text{K.S.}} = 0.242}$	$p_{\text{K.S.}} = 0$
Working 20-30 Hours	$p_{\text{A.D.}} = 0$	$\mathbf{p_{\text{A.D.}} = 0.062}$	–	$p_{\text{K.S.}} = 0$
Working >30 Hours	$p_{\text{A.D.}} = 0$	$p_{\text{A.D.}} = 0$	$p_{\text{A.D.}} = 0$	–

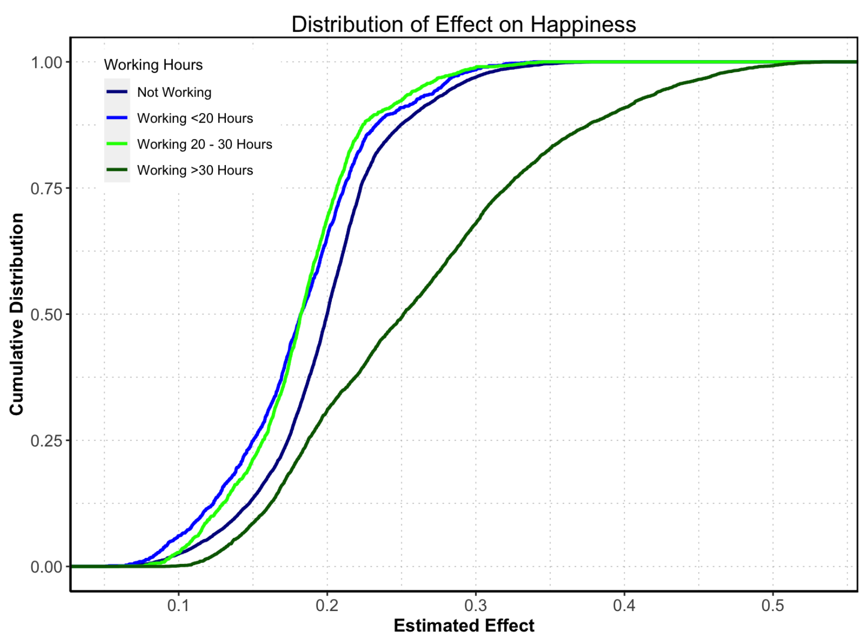

Figure (3.5) **CDF of Medicaid's Effect on Happiness for Working hours.**
This figure shows the distribution of effect for different categories of weekly working hours. The CDF of Effect for those who work more than 30 hours is very different from other categories.

bution functions of effects in different categories of working hours as shown in Figure 3.5. This figure shows that the effect for those who work more than 30 hours per week are very different compared to the effect for other groups. As such, I hereafter isolate the category of individuals who work more than 30 hours per week from others and combine the other three categories, i.e., individuals are divided into two groups: those who work more than 30 hours per week and those who work less than 30 hours per week.

Heterogeneity in Multiple Dimensions. In the previous sections, I looked at the heterogeneity of Medicaid's effect on happiness by age, and weekly working hours. In this section, I show the heterogeneity of Medicaid's effects on happiness with respect to the combination of these important variables.

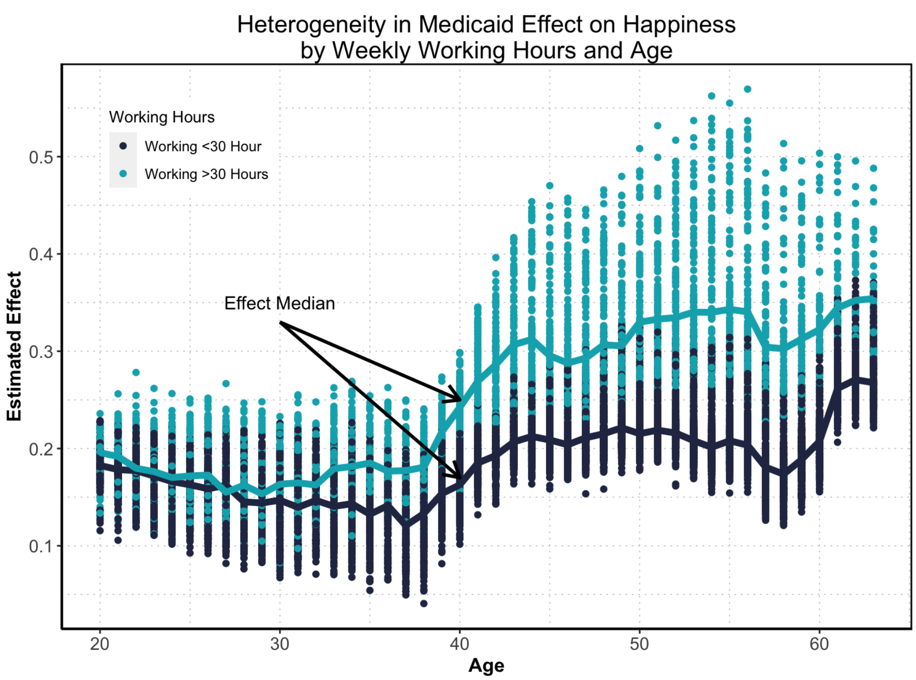

Figure (3.6) **Heterogeneity in Medicaid's effect on Happiness by Age and Working Hours.**
This figure shows how the estimated effect of Medicaid on happiness varies over age, and with respect to weekly working hours. The two colors represent the two levels of working hours (more than 30 hours and less than 30 hours). This figure shows that the Medicaid causes the happiness to be greater for those who are older than 40 years and work more than 30 hours per week than other groups.

Figure 3.6 shows the CLATEs on happiness over the age of individuals and distinguishes the levels of working hours by color. The figure also illustrates the CLATE median for each of the working hours levels over the range of ages. There is a clear distinction between the CLATEs for those who work more than 30 hours compared to the rest. This is particularly true for those older than 40 years.

3.2.4 Potential Mechanisms Behind Medicaid's Effect on Happiness

Besides estimating heterogeneous effect of Medicaid, I try to shed light on a number of mechanisms through which Medicaid may affect happiness. In previous sections I showed that the individual-level effect of Medicaid on happiness is positive. One reason that can potentially explain this result is that with medical insurance coverage, people can attend to their mental and physical health more often and become a happier person. This is shared among all ages and levels of working hours which can explain the observed positive effect for all participants.

Moreover, I show that among older people Medicaid causes those who work more than 30 hours per week to be happier compared to those who work less than 30 hours per week. To explain this heterogeneity in Medicaid's effect, I consider two potential competing mechanisms: "the job lock" and "the pent-up health care demand".

"Job lock" is a situation where senior employees are working full-time and are unable to reduce their working hours, or leave their job and retire; because by doing so they would lose their employer-provided health insurance. If these people had insurance from other sources like Medicaid they would reduce their working hours, or leave the job and retire. In such situation Medicaid would cause those who work more than 30 hours per week to be happier by reducing their working hours. Therefore, I investigate whether people who work more than 30 hours per week before their Medicaid coverage, i.e., in the baseline survey, decrease their working hours after subscribing for Medicaid.

To test whether the job lock mechanism explains the Medicaid's effect or not, I limit my data to those who are above 40 years old and work more than 30 hours per week in the baseline (before subscribing for Medicaid) and perform a two-stage least squares analysis. In this regression, I'm looking for the effect of Medicaid on the follow-up working hours (after subscribing for Medicaid). The results of the regression is reported in Table 3.2 which shows Medicaid has no statistically significant effect on weekly working hours. As such, there is not enough evidence to explain the increase in happiness for those who work more than 30 hours with the job lock.

Table (3.2) **Test of "Job Lock" and "Pent-up Healthcare Demand" Mechanisms**
In this table I show the results of testing two competing mechanisms; job lock and pent-up healthcare demand. I run the following two-stage least square regression:

$$Y_i = \beta_0 + \beta_1 \text{Medicaid}_i + \beta_2 X_i + \varepsilon_i$$

with the following first stage:

$$\text{Medicaid}_i = \alpha_0 + \alpha_1 \text{Lottery}_i + \alpha_2 X_i + \nu_i$$

In the job lock and the pent-up healthcare demand mechanism the Y_i is defined as the working hours and the probability of having a doctor visits respectively in the follow-up survey. To run this regression I limit my data to those people who are older than 40 years of age and work more than 30 hours before their Medicaid coverage. X_i represent all the covariates in Table 2.1.

Outcome	Subsample	Medicaid's Effect
Working Hours	Work>30h, Age>40	0.16 (0.12)
	Work<30h, Age>40	0.02 (0.01)
Doctor Visit	Work>30h, Age>40	0.25^{***} (0.02)
	Work<30h, Age>40	0.17 (0.12)
Note:		*p<0.1; **p<0.05; ***p<0.01

The impact of Medicaid expansion program in Oregon on the labor market was studied in Baicker et al. (2014). The paper considered three measures of labor market activities, namely employment (any earnings), amount of earnings, and whether earnings are above FPL. They did not find any statistically significant effect on labor market activity. The difference between this study and mine is in the considered types of labor market activity measures. My focus is on those individuals who worked more than 30 hours per week before their Medicaid coverage, and the change in the number of working hours afterwards; whereas Baicker et al. (2014) considered the whole population and the above-mentioned measures. Despite the differences, both studies are similarly inconclusive about the effect of Medicaid on labor market activities.

The alternative mechanism for the job lock is the "pent-up health care demand". This mechanism is defined as a spike in demand for health care after obtaining an insurance. The idea is that people without insurance postpone or forego their healthcare needs due to the costs, but after obtaining insurance their utilization for health care increases. Fertig et al. (2016) show that the pen-up demand exists among new enrollees in Minnesota Medicaid plan.

To test this mechanism I limit my data to those people who are older than 40 years of age and work more than 30 hours before their Medicaid coverage. I study the effect of Medicaid on the probability of having a doctor visit, as a measure of health care utilization. My results, presented in Table 3.2, show that those old people who work more than 30 hours per week have higher healthcare utilization after subscribing for Medicaid compared to the other group, i.e., those who are above 40 years old and work less than 30 hours per week. The results indicate that Medicaid increases the health care utilization for the former group, while it has no statistically significant effect on the latter group. Therefore, I conclude that pent-up health care demand exists among older people who work more than 30 hours. As such, this mechanism can explain why this group of people is happier after being covered by Medicaid, since their health needs are satisfied.

3.3 Self-reported Out-of-pocket Medical Costs

The out-of-pocket (oop) medical costs variable takes values 0 or 1, with respect to the answer to "Any out of pocket cost for medical care in the past 6 months?". The positive responses are indicated by the value of 1 and negative responses are indicated by 0. Previous studies such as Finkelstein et al. (2012) show that Medicaid causes the probability of having any out-of-pocket medical costs to decrease by 20 percent. My results also show that this effect is negative for everyone but the size of the effect is different for people with different characteristics. The average of individual effects is -0.22 which is very close to the average effect stated in Finkelstein et al. (2012).

In estimating the effect of Medicaid on out-of-pocket medical costs, the two most important variables in creating effect heterogeneity are age and MSA (people who reside in a

MSA are referred to as urban residents and those who live in a non-MSA are referred to as rural residents).[2]

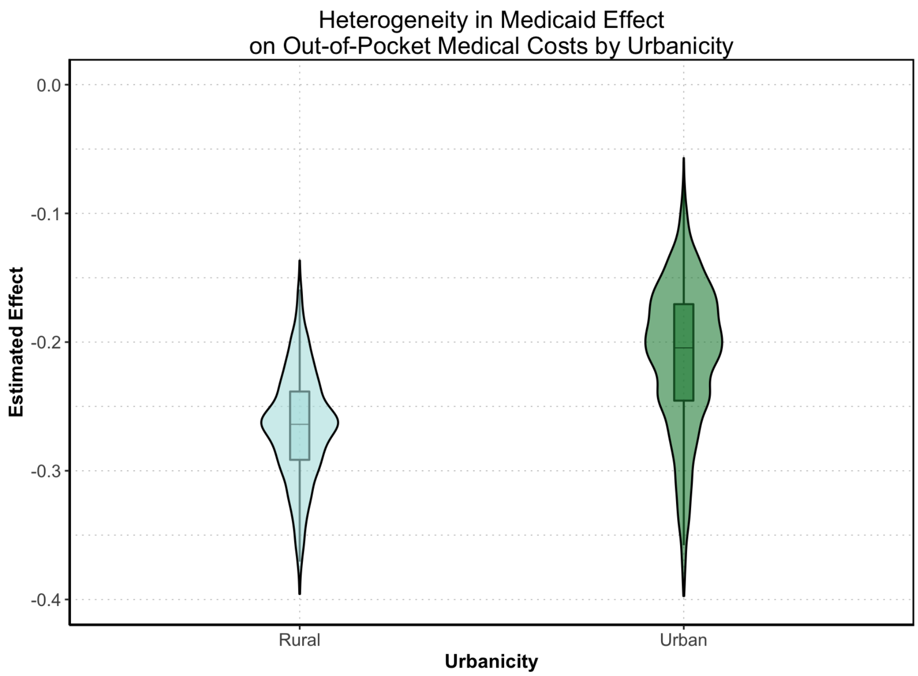

Figure (3.7) **Heterogeneity in Medicaid's Effect on OOP Costs by Urbanicity**
In this figure the distribution of Medicaid's effect on out-of-pocket medical costs is shown with violin plot for those who live in urban and rural areas, separately.

Figure 3.7 shows the distribution of the effect for those who live in urban and rural areas. The *p*-values for the K.S. and A.D. tests are close to zero and confirm that the distributions of effect for these two groups are statistically different. Based on Figure 3.7, the decline in out-of-pocket medical costs is greater for those who live in rural areas compared to those who live in urban areas. This is more visible in Figure 3.8 which shows the heterogeneity in effect of Medicaid with respect to age and MSA, simultaneously. The dashed blue line in Figure 3.8 represents findings by Finkelstein et al. (2012) on the effect of Medicaid on out-of-pocket medical costs which is equal to -0.20. The average of our estimated CLATEs is -0.21 which confirms that my results are aligned with Finkelstein et al. (2012) findings.

[2] For additional results and figures regarding out-of-pocket medical costs, depression, and er visits refer to the Appendix.

My results bring more insight into the effect of Medicaid by showing heterogeneity with respect to individuals' characteristics.

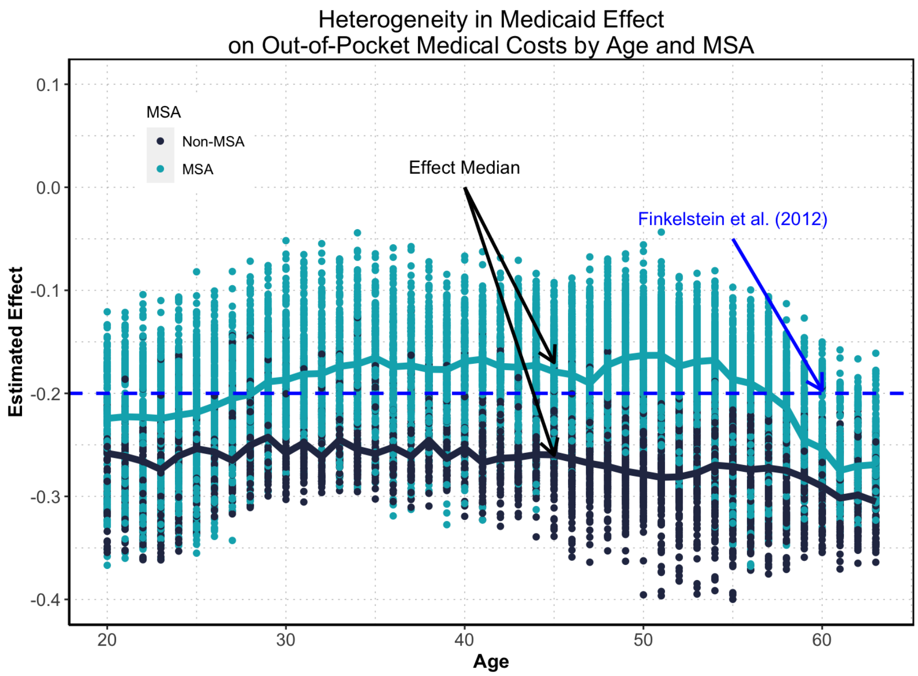

Figure (3.8) **Heterogeneity in Medicaid's Effect on OOP Costs by Age and Urbanicity**
This figure shows how the estimated effect of Medicaid on out-of-pocket costs varies over age, and with respect to MSA. The two colors distinguish the effect for urban and rural individuals. This figure shows that the Medicaid causes the out-of-pocket costs to decline by a larger amount for those who live in rural areas compared to urban areas.

3.3.1 Potential Mechanisms Behind Medicaid's Effect on Out-of-pocket Costs

National Academies of Sciences, Engineering, and Medicine (2018) states that the level of urbanization is an important factor that affects the individual's use of health care services. Previous studies such as the one by National Center for Health Statistics and Center For Disease Control and Preventio (2017) show that those who live in non-metropolitan areas have more doctor visits, emergency room visits, and hospital stays compared to people in metropolitan area.[3]

[3] Residents of central counties of large metropolitan areas and non-metropolitan counties have similarly

Table (3.3) **OOP Costs and Healthcare Utilization Based on Urbanicity.** In this table I compare the out-of-pocket costs and the healthcare utilization of those who live in urban and rural areas. I run the following simple regression

$$Y_i = \beta_0 + \beta_1 \text{MSA}_i + \beta_2 X_i + \varepsilon_i$$

where X_i is the vector of all the covariates in Table 2.1, and the outcome variable is defined as the probability of having out-of-pocket costs and the probability of the doctor visit for each regression. These outcomes are measured before implementing Medicaid expansion. Moreover, I run a two-stage least square regression to compare the effect of Medicaid on the out-of-pocket cost and the doctor visit between rural and urban areas.

Outcome	Subsample	Medicaid's Effect	Baseline Comparison
Out-of-Pocket Costs	MSA	-0.18*** (0.02)	-0.026** (0.00)
	Non-MSA	-0.30*** (0.04)	
Doctor Visit	MSA	0.23*** (0.02)	-0.025** (0.00)
	Non-MSA	0.21*** (0.04)	

Note: *p<0.1; **p<0.05; ***p<0.01

Similarly, using the OHIE dataset, as presernted in Table 3.3, I show that before Medicaid implementation, i.e., in the baseline survey, people who live in non-metropolitan areas have higher probability of doctor visit (as a measure of healthcare utilization) compared to people who live in metropolitan areas. Moreover, in Table 3.3, I show that before implementation of the Medicaid expansion, the probability of having out-of-pocket medical costs is higher among people who live in non-metropolitan areas. This is because these people did not have insurance and their higher utilization of health care is associated with higher out-of-pocket costs. In addition to that, due to lack of competition among providers and the limited resources available in non-metropolitan areas, the out-of-pocket costs for using the health care services is higher. As such, by finding significant differences in baseline service-specific out-of-pocket medical costs between metropolitan and non-metropolitan areas, I propose that the differences in effect of Medicaid on out-of-pocket medical costs could be driven by the disparities in available resources and competition in metropolitan and non-metropolitan areas.

high percentages of residents who lack health insurance.

3.4 Self-reported Depression

The depression variable is the answer to the question "How often have you felt down, depressed, or hopeless over the past two weeks?". This variable takes values from 1 (not at all) to 4 (nearly every day). The estimated effect of Medicaid on depression is negative for everyone which is consistent with those published in Finkelstein et al. (2012) and Baicker et al. (2018).

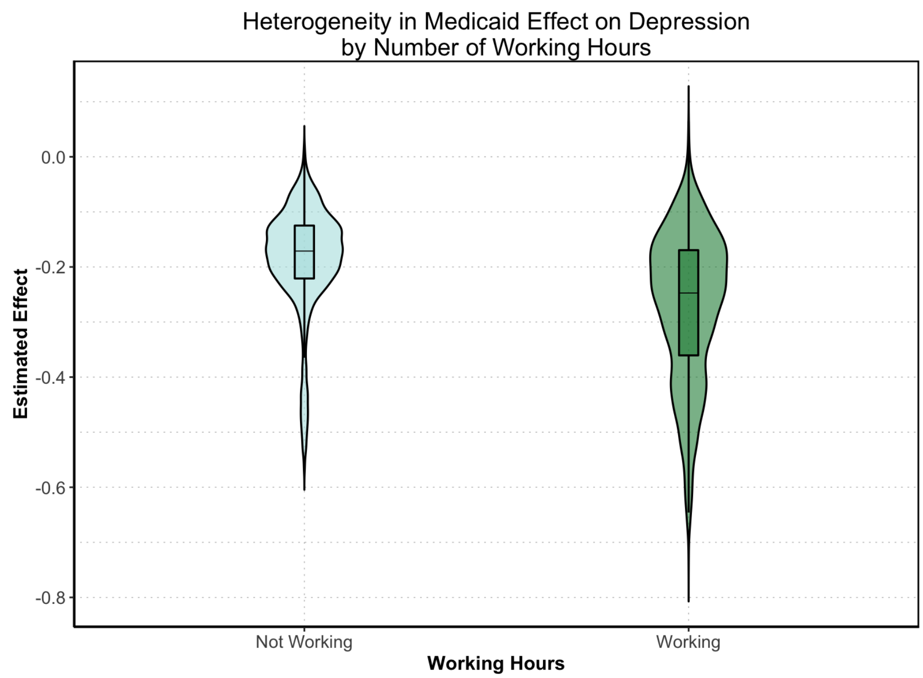

Figure (3.9) **Heterogeneity in Medicaid's Effect on Depression by Work Status**
In this figure the distribution of Medicaid's effect on depression is shown with violin plot for those who do not work and those who work either part-time or full-time, separately.

The most important variables in creating heterogeneity in effect of Medicaid on depression are age and weekly working hours. To investigate the heterogeneity with respect to weekly working hours, I consider two levels of working hours and divide individuals into two groups: those who do not work, and those who work either part-time or full-time.[4]

[4] As shown in Table 2.1 there are four levels of working hours in this dataset: those who do not work, work less than 20 hours, work between 20 and 30 hours, and work more than 30 hours. After observing the

Figure 3.9 shows the distributions of effect for these two levels of working hours. I use the K.S. and A.D. tests to investigate whether the distributions of the effect are statistically different between these two levels. The p-values for these tests are close to zero and confirm that the number of working hours creates heterogeneity in effect of Medicaid on depression. Based on Figure 3.9 Medicaid's effect is stronger in reducing depression for those who work at least several hours per week compared to those who do not work.

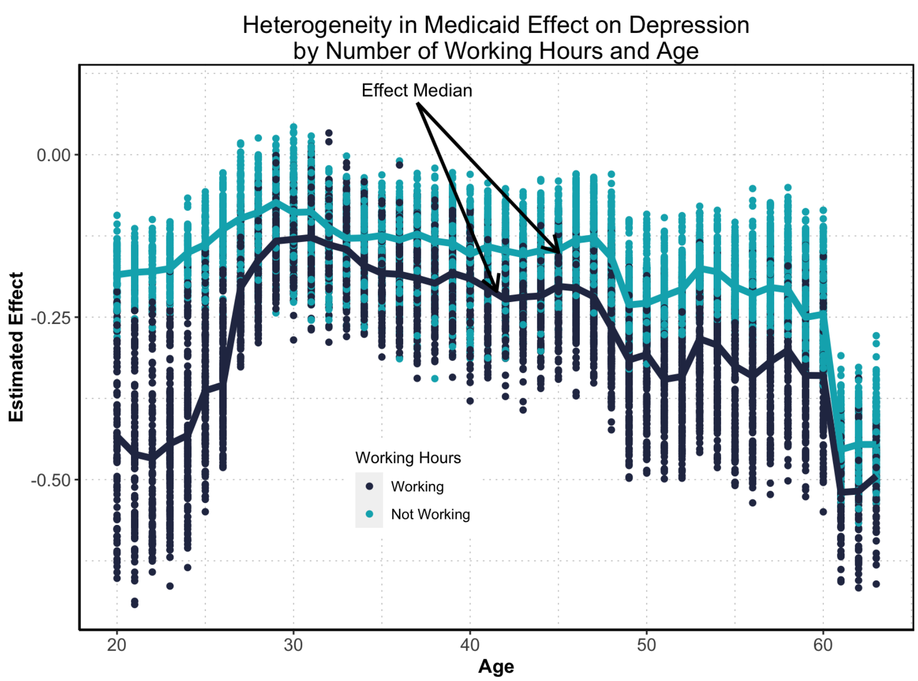

Figure (3.10) **Heterogeneity in Medicaid's Effect on Depression by Age and Working Hours.**
This figure shows how the estimated effect of Medicaid on depression varies over age, and with respect to weekly working hours. The two colors distinguish the effect for those who work and those who do not work. This figure shows that the Medicaid causes the depression to decline by a larger amount for those who work either part-time or full-time compared to those who do not work.

Figure 3.10 shows the heterogeneity in Medicaid's effect in multiple dimensions. The effect median at each age level for both levels of working hours are also shown in Figure 3.10

heterogeneity with respect to working hours and performing K.S. and A.D. test I find that the distribution of effect for the last three groups are not statistically different. As such I merge these three groups and create two groups of working hours: those who do not work and those who work at least several hours per week.

with solid line. This figure shows that the effect varies across different ages and is stronger in reducing the depression for those who work at least several hours per week compared to those who do not work at all, specifically for younger people below 30 years old. The difference in the size of the effect is evident among the younger people. In order to explore the driving factors of this heterogeneity I focus on the characteristics of individuals and use the previous findings in the literature on depression.

3.4.1 Potential Mechanisms Behind Medicaid's Effect on Depression

McGee and Thompson (2015) study the relationship between unemployment and depression among young adults. They show that the risk of depression is higher among those who do not work. In addition to that, Fletcher and Frisvold (2009) finds that individuals with higher education use more preventive medical care.

Table (3.4) **Depression and Education Among Young Adults.**
In this table I compare the Depression and the education level of young adults who work and who do not work. I run the following simple regression

$$Y_i = \beta_0 + \beta_1 \text{Work}_i + \beta_2 X_i + \varepsilon_i$$

where X_i is the vector of all the covariates in Table 2.1, and the outcome variable is defined as the probability of being diagnosed with depression for the regression. This outcome is measured before implementing Medicaid expansion. Moreover, I run a two-stage least square regression to compare the effect of Medicaid on Depression among young adults who work and do not work. The baseline comparison columns report the average of depression for those who work minus that of those who do not work. In this table I also compare the proportion of having high school diploma or higher degree for these two groups of people.

Outcome	Subsample	Medicaid's Effect	Baseline Comparison
Depression	age<30, Work	-0.59^{**} (0.21)	-0.09^{***} (0.01)
	age<30, Do Not Work	-0.16^{*} (0.16)	
Education	age<30, Work	-	0.32
	age<30, Do Not Work	-	0.22
Note:			*p<0.1; **p<0.05; ***p<0.01

Using my data I can confirm these findings, as shown in Table 3.4, and provide an explanation for the heterogeneity in effect of Medicaid on depression among young adults.

I find that in the baseline survey, i.e., before Medicaid coverage, among the young adults who are less than 30 years old, those who are unemployed are more depressed compared to those who work at least for several hours per week. Figure 3.10 shows that Medicaid has a greater effect on those who work compared to those who do not work. According to the literature, this difference could be associated to the difference in the individuals' level of education. Using my data, I confirm that among the young adults, a higher proportion of those who work have high school diploma or higher, compared to those who do not work. Using this finding I conclude that educated people benefit more from Medicaid and improve their mental health.

Chapter 4

PREDICTING CAUSAL EFFECTS IN AN UNTREATED POPULATION BY GENERALIZING ESTIMATED CAUSAL EFFECTS

4.1 Introduction

The goal of this chapter is to study the relevance of experiment results, obtained and presented in the previous chapters, to a particular population where the intervention under study is not yet implemented.

In a Randomized Control Trial (RCT) study, where the intervention is randomly assigned to the participants, the estimated effect of the treatment is unbiased. This estimated effect is internally valid for the sample as it is an asymptotically unbiased estimate of the parameter of the interest within the sample. Internal validity refers to the degree to which we can trust our estimate of the effect in the population under study.

The estimated effect is readily generalizable to the larger population, from which the participants of the study are selected, only if the selection process is also randomized. For example, consider a population consisting of people with characteristics A, B, and C. The results of an RCT study over a sample from this population in which the sample lacks any participant with the C characteristic, is not necessarily generalizable to the whole population. If the selection process for such RCT study is also randomized, the unbiased estimated effect over the sample is generalizable to the whole population. In this case it is said that the estimated effect is externally valid for the larger population as it is an asymptotically unbiased estimate of the effect for the larger population. External validity refers to the extent to which the experiment's results can be generalized to other populations (Weisberg et al., 2009; Greenhouse et al., 2008; Shadish et al., 2002).

External validity can generally be divided into two broad categories, generalizability and transportability. Generalizability refers to the extent the internally valid estimate of the effect in a sample can be representative of the encompassing population. In contrast, transportablity refers to the extent the internally valid estimate of the effect in a study population can be representative of a distinct target population (see Figure 4.1).

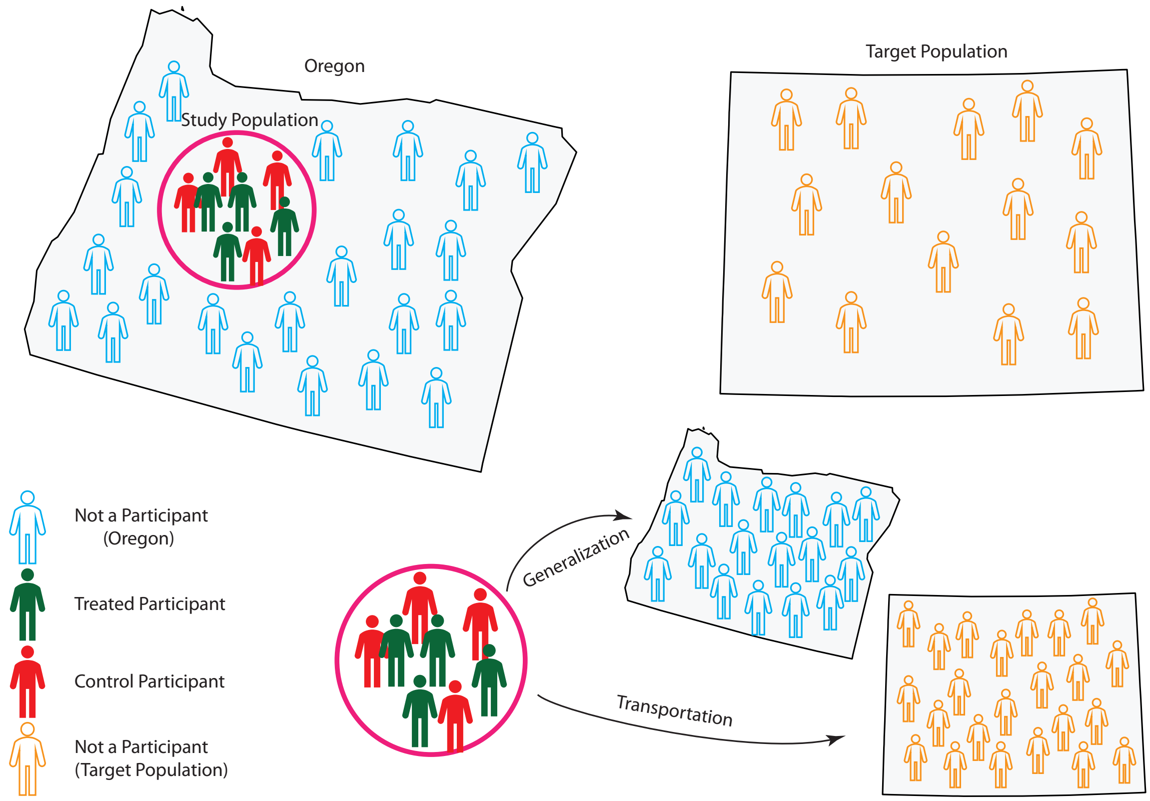

Figure (4.1) **Schematics of generalizability and transportability.** This figure shows the difference between generalizability and transportability. In the state of Oregon a group of people were selected and the Oregon health insurance experiment was conducted among them. The policy makers want to extend the results from this experiment beyond this specific group. If the results extended to the population that these people were sampled from, the process is called generalizing the experiment results. Whereas if the results extended to another population without any overlap with people in the experiment, it is called transporting results from the study population to the target population.

In observational studies, in which the intervention is not randomly assigned to the participants, internal validity is usually achieved using causal inference methods like Instrumental Variable. In such methods an unbiased estimate of the causal effect for the intervention is obtained. In this chapter, I would like to study the transportability of the effect of the Medicaid expansion program to populations where this program is yet to be implemented.

Transporting the estimated effect of Medicaid expansion program implemented in Oregon requires both internal and external validity. Internal validity of the estimate is the subject of Chapter 3. External validity of this estimate for transporting results to the population of another state is discussed here. Given the inherent differences between the residents of Oregon compared to the state where the results are to be transported, there exists an external validity bias. I need to address this bias and adjust the differences in order to transport Medicaid expansion program's effect (Flay et al., 2005). In this chapter, I use several statistical methods to transport these estimated effects to another population, hereafter refered to as target population.

The problems of generalizability and transportability have been addressed by several studies in the literature (Cole and Stuart, 2010; Stuart et al., 2011; Haneuse et al., 2009; Stuart and Rhodes, 2017; Tipton, 2013; Lesko et al., 2017; Pearl and Bareinboim, 2011; Buchanan et al., 2018; Degtiar and Rose, 2021). Most of these studies use weighting-based approaches similar to Horvitz–Thompson (Horvitz and Thompson, 1952) weighting to estimate the average effect in the target population.

Propensity-score-based methods are frequently used in quasi-experiments and observational studies to address treatment selection biases (Rosenbaum and Rubin, 1983). The weighting methods based on propensity score for generalization and transportability are a modification to these models. The general idea behind weighting methods is to make the study and the target populations resemble each other. One of such weighting methods defines the propensity score as a model of experiment participation probability conditional on observable covariates. This propensity score is then used to address differences between the study population and the target population.

Stuart et al. (2011) proposed using Inverse Propensity Weighting (IPW) to measure the similarity of individuals in the study population and the target population. Lesko et al. (2017) used g-formula and IPW estimators for generalizing results from a study population to a target population. Kern et al. (2016) examined different Machine Learning algorithms for predicting weights used in the weighting methods. They also used the Bayesian Additive Regression Trees (BART) method to predict the effect in the target population.

In this chapter, my focus is on the common weighting method, specifically IPW. I use the individual-level data on treatment, instrumental variable, outcome, and baseline covariates from the Oregon population to estimate Conditional Local Average Treatment Effects (CLATEs) in Oregon experiment using Generalized Random Forest (GRF)[1]. Moreover, I assume that the individual-level data on baseline covariates from the target population are available. Using the baseline covariates, I estimate the propensity scores ($P(\mathbf{X}) = Pr(S = 1|\mathbf{X})$). Under some general assumptions, I estimate the average effect in the target population using estimated CLATEs and the weighted average effect where the weights are inverse of estimated propensity scores. In addition to the inverse propensity score weighting strategy I use the GRF model which is trained on the Oregon data to predict the effects in the target population.

I use several methods to estimate the propensity scores, namely a linear method with Logistic Regression (LR), and more flexible non-linear Machine Learning methods including Random Forest (RF), and Gradient Boosting Method (GBM). I compare the results obtained from all these methods on a synthetic set of covariates which I generate using a Copula strategy.

4.2 Notation and Assumptions

I follow the potential outcome framework for transporting the effects from the study population to the target population. Consider population $\mathcal{P}_1$ with a set of N independent identically distributed individuals in the study population (individuals participating in Oregon experiment). Variables $\mathbf{X}_i, Y_i, D_i, Z_i$, where $i = 1 \cdots N$, are observable in the study population. $\mathbf{X}_i$ is the covariates vector including age, education, gender, income level, etc. . D_i is the treatment variable and $D_i = 1$ if an individual is assigned to the treatment, meaning enrolled in the Medicaid program in the context of the Oregon experiment. Y_i is the outcome variable, which is the happiness outcome in the Oregon experiment considered herein. $Y_i(1)$ represents the potential outcome under treatment, and $Y_i(0)$ represents the potential outcome under untreated conditions. According to the fundamental problem of causal inference, both potential outcomes can never be observed for any individual (Holland, 1986), and at most one of the outcomes can be observed at any time, i.e. $Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$.

[1] Please refer to the Chapter 1 for more details.

Since the Oregon experiment is not a randomized experiment and they used a lottery drawing to assign individuals to the treated and untreated groups, I consider lottery status of each individuals as the instrumental variable denoted by Z_i.[2]

Now consider $\mathcal{P}_2$ with M individuals to be the population for which the estimate of the treatment effect for its individuals is of interest. This population can, for example, be any of the states in the U.S. that has not implemented the Medicaid expansion program yet, e.g. Florida. I use the term "Target Population" for $\mathcal{P}_2$ hereafter. Note that $Y(0)$, $Y(1)$, D, and Z are not observed for the individuals in the target population, but it is assumed that all statistics about the distribution of covariates in this population are available.

Let's define an indicator S for the population membership. I denote individuals in the study population ($\mathcal{P}_1$) with $S = 1$, and individuals in the target population ($\mathcal{P}_2$) with $S = 0$. By defining the individual treatment effect as $\tau_i = Y_i(1) - Y_i(0)$, the Average Treatment Effect (ATE) in the study population is $\tau_{\text{OR}} = E[\tau_i = Y_i(1) - Y_i(0)|S = 1]$, and the ATE in the target population is $\tau_{\text{TP}} = E[\tau_i = Y_i(1) - Y_i(0)|S = 0]$.

Under two specific conditions the ATE in the study population and the target population would be identical: 1. If the treatment effect is constant, i.e. $\tau_i = C$ where C is a constant, 2. If individuals in the study population and target population are randomly and independently sampled from another population. In this study neither is true. As shown in Chapter 3 the treatment effect is heterogeneous. Additionally, the target population is not a randomly selected sample. As such, I need to consider the differences between individuals in terms of their characteristics in both populations and adjust these differences.

4.2.1 Propensity Score

The propensity score is defined as the probability of treatment assignment conditional on the baseline covariates. This definition of propensity score is commonly used in randomized experiments and observational studies. In randomized experiments the true value of the propensity scores are known and are prescribed by the experimenter depending on the experiment design. In observational studies, however, the propensity scores are not known

---

[2]Refer to Chapter 1 for detailed explanation of the Oregon Health Insurance Experiment.

and should be estimated given the observable covariates.

For the purpose of transporting CLATEs from the study population to the target population, propensity scores are used to match individuals in the study population to individuals in the target population (Cole and Stuart, 2010). In this context, the formal definition of propensity scores is

$$P(\mathbf{X}) = Pr(S = 1|\mathbf{X}) \tag{4.1}$$

where S indicates membership in the study population.

Propensity score is defined as the predicted probability for individuals to be a member of the study population conditional on the observed covariats. The propensity score quantifies the differences between individuals in the study population and target population and can be used as a balancing score (Rosenbaum and Rubin, 1983). A balancing score is any function $b(\mathbf{X})$, such that $\mathbf{X} \perp S|b(\mathbf{X})$. In other words, conditional on $b(\mathbf{X})$ the distribution of $\mathbf{X}$ is independent of population membership S. With this definition the individuals in $\mathcal{P}_1$ and $\mathcal{P}_2$ with the same balancing score can be matched. In the remainder of this study, I use propensity scores as the balancing score between study and target population.

4.2.2 Assumptions

To use propensity scores for transporting results from the study population to the target population, the following assumptions need to be made:

1. Stable Unit Treatment Value Assumptions (SUTVA). This is one of the common assumptions made when estimating the treatment effect in the study population (Imbens and Rubin, 2015). Following Tipton (2013), transportability of the treatment effect requires defining two versions of SUTVA, one for the study population, SUTVA($\mathcal{P}_1$), and one for the target population, SUTVA($\mathcal{P}_2$).

SUTVA($\mathcal{P}_1$) states that if the same treatment is given to the individual i in the study population, the response of the individual would be the same irrespective of how other individuals are assigned to the treatment or how they respond. In other words, SUTVA($\mathcal{P}_1$) states that for the individual i in the study population there is only one version of each treatment level available, and the potential outcome for the individual does not vary with the treatments assigned to other units. Formally, $\forall i \in \mathcal{P}_1$ if $D_i = D_i'$ then $Y_i(D_i) = Y_i(D_i')$.

SUTVA($\mathcal{P}_2$) is defined in relation to treatment effect in the the study population and the target population. Let's define the treatment effect for individual i in the study population as $\tau_i(S = 1) = Y_i(D = 1, S = 1) - Y_i(D = 0, S = 1)$ and in the target population as $\tau_i(S = 0) = Y_i(D = 1, S = 0) - Y_i(D = 0, S = 0)$. In parallel to the assumption for SUTVA($\mathcal{P}_1$) for individual i, SUTVA($\mathcal{P}_2$) states that the treatment effect would be the same regardless of which population individual i is selected from. That is, for a pair of membership variable S_i and S_i', if $S_i = S_i'$ then $\tau_i(S_i) = \tau_i(S_i')$.

2. Conditional exchangeability over S for populations. This assumption states that the individual treatment effects are independent of individuals' membership to either $\mathcal{P}_1$ or $\mathcal{P}_2$ conditional on propensity score $P(\mathbf{X})$.

$$\tau_i = Y_i(1) - Y_i(0) \perp S_i | P(\mathbf{X}) \tag{4.2}$$

For this assumption to hold it is necessary to include those covariates which generate heterogeneity in the effect and are important in explaining the differences between the two populations,i.e., $\mathcal{P}_1$ and $\mathcal{P}_2$, in the covariates vector $\mathbf{X}$. It is also as important to not include the indicator of the population membership (S) in the covariates vector.

3. Positivity of population participation assumption. This assumption states that each individual has positive probability to be included in the study population. As such, the estimated propensity score needs to be between 0 and 1.

$$0 < Pr(S = 1|\mathbf{X}) < 1 \tag{4.3}$$

Assumptions 2 and 3 are collectively called **Strong Ignorability** assumption.

4. Compliance score is the same in both populations. This assumption states that if we had a chance to conduct the same experiment as done in the study population in the target population, the rate of compliers would be the same in both the study population and the target population.[3]

5. Individuals with the same covariates have the same CLATE in both populations. Suppose that individual i with $S_i = 0$ and individual j with $S_j = 1$ have the same characteristic covariates vector, i.e., $\mathbf{X}_i = \mathbf{X}_j$. This assumption states that the CLATE for these two individuals should be equal regardless of their membership status, i.e., $\tau_i(\mathbf{X}_i, S_i) = \tau_j(\mathbf{X}_j, S_j)$.

6. All differences between the study population and the target population can be explained given the observable covariates. Since propensity scores are estimated based on observable covariates in the model and they are used for balancing the two populations, it is important to assume that there are no unobserved variables that can explain the differences between study population and target population.

4.3 Statistical Method

In Chapter 3, I estimated the Conditional Local Average Treatment Effects (CLATEs) of Medicaid on happiness using the Generalized Random Forest (GRF) method, for each individual in the Oregon experiment. Given these estimated effects and assumptions listed in the previous section, I can estimate the propensity scores and use them to calculate several estimators which estimate the average treatment effect in the target population. In the following section, I describe a few methods for estimating the propensity scores and discuss several methods for estimating the average treatment effects in the target population.

4.3.1 Estimating the probability of participation in an experiment, Propensity Score

In transporting results from the study population to the target population, I define the propensity score as the probability of being a participant of the study population and use it to balance the covariates in the study and the target populations. In other words, the propensity score is used to adjust for the differences between the study and the target population.

[3] Compliance score is the predicted probability that each person complies with the treatment assignment and is estimated as $P_{C_i}(x) = E[D_i|Z_i = 1, X_i = x] - E[D_i|Z_i = 0, X_i = x]$

Estimating the propensity scores can be done using a variety of approaches. Here I focus on three methods. First, I use the Logistic Regression (LR) method which is the most common method for estimating the propensity score. LR uses the linear combination of covariates to predict the propensity scores. Second, I use the Random Forest (RF) algorithm which is an ensemble machine learning method. It considers a non-linear combinations of covariates and predict the propensity scores (Breiman, 2001). The third model I use here, is the Gradient Boosting Model (GBM) which combines decision trees with boosting algorithms. That is, GBM models build trees sequentially, where each new tree helps to correct errors made by previously trained tree. RF and GBM are two flexible models for estimating the propensity scores and this is their advantage over LR. In LR model, it is important to know the correct form of the model. Misspecification of the model in the LR method may lead to a poor performance in estimating the propensity scores.

4.3.2 Estimating Medicaid Effects in the Target Population

There are several approaches for estimating the average effect of Medicaid in the target population. In the following I discuss a weighting-based method that uses the estimated propensity scores to adjust the differences between the study population and the target population. Another method that I use, is based on predicting the effects in the target population using the trained GRF model over the study population data.

Inverse odds weighting. This method proposed by Cole and Stuart (2010) and it weights experiment participants using the inverse odds. The goal of inverse odds weighting is to use the propensity score and make the individuals in the experiment similar to individuals in the target population and match the resulting effect for that individual. The weights are define as follows:

$$W_i = \begin{cases} \frac{1-\hat{P}_i(\mathbf{X})}{\hat{P}_i(\mathbf{X})} & S_i = 1 \\ 0, & S_i = 0 \end{cases}$$

Since individuals in both populations ($\mathcal{P}_1$ and $\mathcal{P}_2$) are not randomly selected, the propensity scores ($\hat{P}_i$) are estimated using LR, RF, and GBM and used in weights.

Estimating treatment effect in the target population using GRF trained on study population data. Another method for transporting results from study population to the target population is using the trained GRF model on the study population data to predict the effect in the target population. In Chapter 3, I used the treatment variable information, instrumental variable, outcome, and covariate vectors for each individual in the Oregon experiment to train the GRF model. The data were sourced from the Oregon experiment and the estimated treatment effect for each individual was predicted using out-of-bag prediction method. That is, for estimating the Medicaid effect for individual i, the algorithm identifies and uses trees that did not use this individual in their training process to estimate the treatment effect for this individual.

For transporting these effects to another population, I use the trained GRF and individuals' characteristics in the target population to predict the effect in the target population. The results can be interpreted as the average effect of Medicaid on individuals' happiness in the target population if they had access to the same Medicaid insurance as in Oregon.

4.4 Data

For demonstrating method described in the previous section I use a set of synthetic target population and the Oregon Health Insurance Experiment (OHIE) dataset as the study population. The following sections describe the details of the study population as well as the steps for generating the synthetic target population.

4.4.1 Study Population

The study population used here is the OHIE dataset introduced in Chapter 2. I estimated the heterogeneous impact of Medicaid enrollment on individuals' happiness in Oregon as described in Chapter 3. I showed that among all variables *age*, *weekly working hours*, and *income* are the three most important variables such that they contribute the most in dispersion of Medicaid's effect. In what follows I'll focus only on these three variables. In Oregon experiment the average of individuals' age is 43 years old with a standard deviation of 12.1. About 52.1% of individuals do not work, 9.6% work less than 20 hours per week, 11.1% work between 20-30 hours per week, and 26.2% work more than 30 hours per week. In this experiment a majority of people are very poor. About 36.5% have income below 50% of the Federal Poverty Level (FPL), 11.1% have income between 50 to 75% of FPL, 13.5% have income between 75-100% FPL, 15.6% have income between 100-150% FPL, and 10.5% have income above 150% FPL.

4.4.2 Target Population

I simulate a target population with N = 15,000 individuals. I assume that the only information available from the target population is the distribution of covariates and the outcome variable is not observed. I consider the same covariates for the target population as study population with different distributions. To simulate correlated random variables, I use a Normal Gaussian Copula. Using Copula I generate the covariates in the target population such that the correlation matrix in the target population is very similar to the correlation matrix in study population but the covariates' marginal distributions are different.

I study the average effect in the target population under different scenarios where the distribution of important covariates are different between study population and target population.

Although the propensity score is used to balance the covariates between the study and the target populations, in some cases imbalance in baseline characteristics may persist if the statistical model used to calculate the propensity score is mis-specified. The most commonly used statistic to assess the balance of covariate distributions between two populations after weighting with propensity scores is the Absolute Standardized Mean Difference (ASMD). Typically, a ASMD greater than 0.1 can be considered as a sign of covariate imbalance (Zhang et al., 2019). I calculate the ASMD for covariates before weighting and after weighting to show that for covariates that I change the distribution in the target population the ASMD is large and after weighting by propensity scores the ASMD decreases to below 0.1. In other words, two population become balanced.

4.5 Results

The study population, $\mathcal{P}_1$, is different from the target population, $\mathcal{P}_2$, on its baseline characteristics, specifically those characteristics that contribute the most in the disparity of the effect.

To demonstrate the methods for transporting results from $\mathcal{P}_1$ to $\mathcal{P}_2$, I change only the distribution of one covariate in $\mathcal{P}_2$ at a time and keep the distribution of other variables the same as those in $\mathcal{P}_1$. I kept the correlation matrix between all covariates of interest the same between the study population and the generated target population in case 1 to case 5. In the last case I change the correlation between covariates but kept the covariate distributions the same. I then show the average effect of Medicaid expansion program for people in the target population if the population had access to the same Medicaid insurance as in $\mathcal{P}_1$.

Case 1: Changing only the distribution of Age. Figure 4.2 shows the distribution of age, working hours, and income in the Oregon experiment and the target populations. The distribution of age in the target population is selected such that individuals are younger compared to individuals in the Oregon experiment. In addition to that this figure shows that the proportion of people in different categories of working hours and income are approximately the same. Other variables' distributions are also kept the same.

Table (4.1) **Case 1. Covariate means in study and target populations and Absolute Standardized Mean Difference (ASMD) before and after weighting.** This table shows the distribution of covariates in the study and target populations based on different levels of covariates. It also shows the (ASMD) of covariates before weighting and after weighting by propensity scores.

Variable	Level	Study population	Target population	ASMD	ASMD (after weighting)		
				(before weighting)	Logistic	Random Forest	GBM
Age		43	32.8	0.99	0.09	0.015	0.034
Employment	Do not currently work	52.1%	52.1%	0.017	0.008	0.013	0.045
	Work < 20 hours per week	9.6%	9.6%	0.002	0.006	0.004	0.009
	Work 20-29 hours per week	11.1%	11.1%	0.015	0.009	0.003	0.022
	Work 30+ hours per week	27.2%	27.2%	0.009	0.001	0.015	0.041
Income (% FPL)	<50%	36.5%	36.5%	0.022	0.023	0.042	0.011
	50-75%	11.1%	11.2%	0.004	0.004	0.003	0.030
	75-100%	13.5%	13.6%	0.014	0.008	0.008	0.006
	100-150%	15.6%	15.9%	0.011	0.008	0.003	0.021
	Above 150%	10.5%	10.4%	0.004	0.005	0.034	0.020
Gender	Female	60.7%	60.7%	0.001	0.004	0.045	0.032

Although the propensity score is used to balance the covariates between the study and the target populations, in some cases imbalance in baseline characteristics may persist if the statistical model used to calculate the propensity score is misspecified. The most commonly used statistic to assess the balance of covariate distributions between two populations after weighting with propensity scores is the Absolute Standardized Mean Difference (ASMD). Typically, a ASMD greater than 0.1 can be considered as a sign of covariate imbalance (Zhang et al., 2019). I calculate the ASMD for covariates before weighting and after weighting to show that for covariates that I change the distribution in the target population the ASMD is large and after weighting by propensity scores the ASMD decreases to below 0.1. In other words, two population become balanced.

Table 4.1 shows the distribution of covariates in the study population and the target population for the case that the distribution of age is different between two populations. The table also shows the ASMD of covariates between the study population and the target population before and after weighting with propensity scores estimated using the three different approaches LR, RF, and GBM. As shown in the table the ASMD before weighting is greater than 0.1 for age. The ASMD for this covariate drops after weighting with the

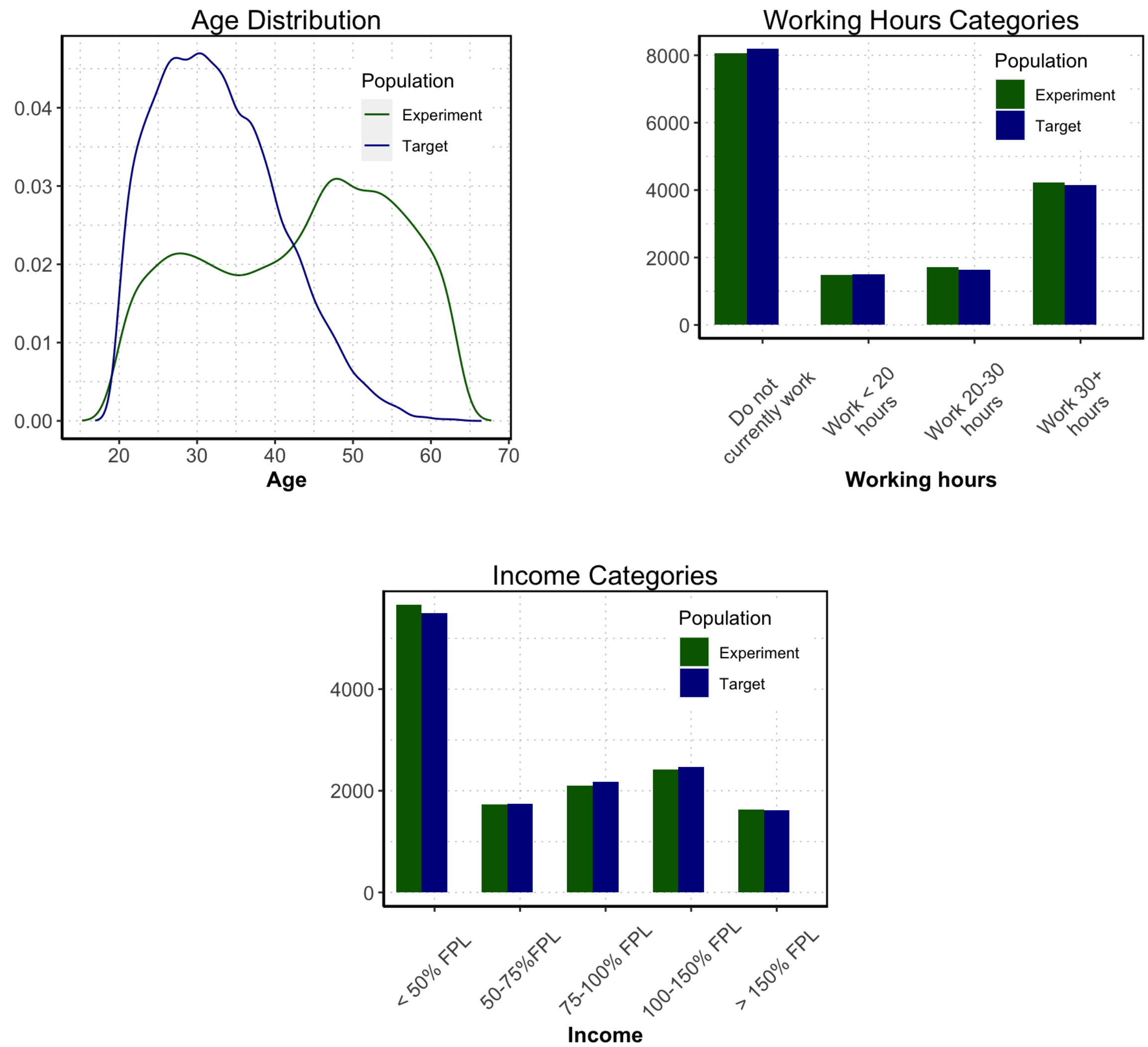

Figure (4.2) **Case 1. Distributional shift of age.**
This figure shows the distribution of age, weekly working hours, and income in the study population (green) and target population (blue). In this case the only difference between the study population and the target population is the distribution of age. In the target population a majority of people are younger (around 30 years old), but the distribution of weekly working hours, income, and other variables (not shown here) are the same.

odds ratio predicted with any of the propensity score estimation methods.

Table (4.2) **Average effects in the study population and target population.** This table shows the estimated average effects based on different models for the study population and target population. The first column in this table shows different cases in generating data for the target population. The second column represent the average effect in study population using two-stage least squares (2SLS) method. The third column shows average of estimated effects in the study population using GRF (from Chapter 3). In the fourth column I use the trained GRF on study population data and predict the average effect for the target population. The last three columns shows the average effect in the target population using weighting method where the weights are estimated with Logistic Regression, Random Forest, and Gradient Boosting Model.

Case	2SLS	GRF	GRF	Weighting Methods (Target population)		
	Study population	Study population	Target population	Logistic	Random Forest	GBM
1. Age only	0.202	0.214	0.179	0.181	0.190	0.179
2. Working Hours only	0.202	0.214	0.211	0.222	0.222	0.212
3. Income only	0.202	0.214	0.219	0.216	0.211	0.218
4. Age, Working Hours, Income	0.202	0.214	0.174	0.181	0.197	0.177
5. Age and Working Hours	0.202	0.214	0.259	0.265	0.255	0.241
6. Gender only	0.202	0.214	0.218	0.216	0.214	0.216
7. Correlation matrix changed	0.202	0.214	0.218	0.214	0.214	0.214

For populations with these characteristics the average effect computed with different methods are reported in Table 4.2. Also shown in this table is the LATE parameter estimated using Two-Stage Least-Squares model in Oregon experiment which is 0.202. This number is very close to the arithmetic average effect in Oregon estimated using GRF.

Table 4.2 also shows the average effect of Medicaid enrollment in the target population using GRF which is 0.179. The lower average effect compared to Oregon experiment is expected as the simulated target population is younger than the study population and as I showed in Chapter 3, the effect of Medicaid enrollment on happiness is smaller for younger people than older people.

The average treatment effects in the target population estimated using three different weighting models are also shown in the table. These three models differ only in the way propensity score is calculated, namely, Logistic Regression (LR), Gradient Boosting Model (GBM), and Random Forest (RF).

Case 2: Changing only the working hours distribution. Figure 4.3 shows another case where the distributions of age, income and other variables are the same between the two populations but the distribution of working hours is different. As shown in Figure 4.3, the working hours plot, the proportion of individuals in the first three categories of working hours is very different between two populations but the proportion of those who work more than 30 hours per week are not very different. This can be seen in the Table 4.3 under the ASMD column before weighting. The large ASMDs for the first three groups of working hours show that the distributions of these groups are different but the distribution of those who work more than 30 hours per week are not very different between two populations. After estimating propensity scores, with all three methods, and weighting, the two populations become balanced.

Table (4.3) **Case 2. Covariate means in study and target populations and Absolute Standardized Mean Difference (ASMD) before and after weighting.** This table shows the distribution of covariates in the study and target populations based on different levels of covariates. It also shows the (ASMD) of covariates before weighting and after weighting by propensity scores.

Variable	Level	Study population	Target population	ASMD (before weighting)	ASMD (after weighting)		
					Logistic	Random Forest	GBM
Age		43	43	0.006	0.014	0.012	0.012
Employment	Do not currently work	52.1%	30.9%	0.439	0.001	0.012	0.019
	Work < 20 hours per week	9.6%	19.8%	0.292	0.002	0.021	0.014
	Work 20-29 hours per week	11.1%	24.3%	0.355	0.001	0.022	0.017
	Work 30+ hours per week	27.2%	24.8%	0.056	0.002	0.028	0.010
Income (% FPL)	<50%	36.5%	36.0%	0.022	0.009	0.012	0.003
	50-75%	11.1%	11.2%	0.004	0.014	0.020	0.011
	75-100%	13.5%	13.6%	0.014	0.004	0.010	0.012
	100-150%	15.6%	15.9%	0.011	0.006	0.001	0.001
	Above 150%	10.5%	10.4%	0.004	0.011	0.003	0.010
Gender	Female	60.7%	60.7%	0.001	0.029	0.032	0.006

The average effects for this case are reported in Table 4.2. The average effect for the target population using GRF prediction and the three weighting methods are all around 0.211 which is very similar to the average effect in Oregon. This is expected considering the results shown in Chapter 3. According to the results presented in Chapter 3, the first three categories of the working hours have almost identical effects. That is, the effect is not heterogeneous with respect to these three categories. This is while the effect is very

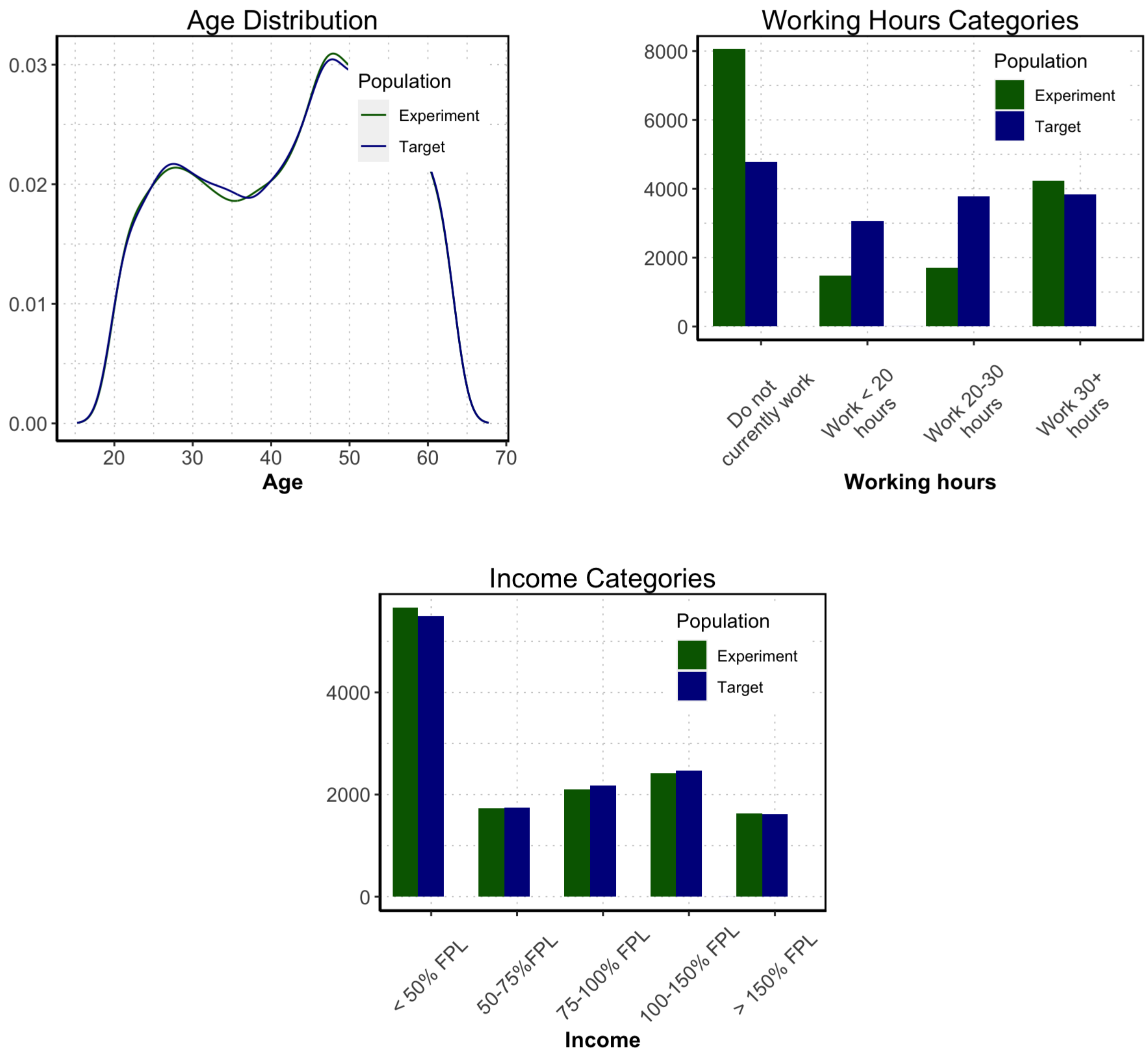

Figure (4.3) **Case 2. Distributional shift of weekly working hours.**
This figure shows the distribution of age, weekly working hours, and income in the study population (green) and target population (blue). In this case the only difference between study population and target population is the distribution of weekly working hours. In the target population the proportion of those who do not work, those who work less than 20 hours, and those who work between 20 to 30 hours are higher than that of the study population. The distribution of age, income, and other variables (which are not shown here) are the same.

different for those who work more than 30 hours per week. As such, since the proportion of those who work more than 30 hours per week is similar between these two populations, it is expected that the effect in the target population be similar to the effect in Oregon experiment.

Case 3: Changing only the income distribution. In this case the distribution of age, working hours, and other variables are the same between the two populations, but income has different distributions in the study population and the target population. Figure 4.4 shows these distributions. The generated data for the income variable in the target population is such that for all categories of income the distributions are different. Table 4.4, shows the proportion of individuals' income in the study population and the target population in columns 3 and 4. In addition to that, in the column ASMD before weighting, for all income categories, all ASMDs are greater than 0.1. It confirms that the the distribution of income is very different betwen two populations. Table 4.4 also show that the ASMD after weighting under all three propensity score estimation methods are very small and can conclude that the differences between two populations are adjusted after weighting.

Table (4.4) **Case 3. Covariate means in study and target populations and Absolute Standardized Mean Difference (ASMD) before and after weighting.** This table shows the distribution of covariates in the study and target populations based on different levels of covariates. It also shows the (ASMD) of covariates before weighting and after weighting by propensity scores.

Variable	Level	Study population	Target population	ASMD	ASMD (after weighting)		
				(before weighting)	Logistic	Random Forest	GBM
Age		43	43	0.006	0.006	0.012	0.015
Employment	Do not currently work	52.1%	52.1%	0.017	0.011	0.014	0.010
	Work < 20 hours per week	9.6%	9.6%	0.002	0.008	0.013	0.013
	Work 20-29 hours per week	11.1%	11.1%	0.015	0.001	0.005	0.008
	Work 30+ hours per week	27.2%	27.2%	0.009	0.017	0.021	0.015
Income (% FPL)	<50%	36.5%	19.5%	0.386	0.008	0.002	0.016
	50-75%	11.1%	19.6%	0.236	0.004	0.033	0.024
	75-100%	13.5%	10.0%	0.111	0.002	0.035	0.003
	100-150%	15.6%	20.4%	0.125	0.004	0.018	0.022
	Above 150%	10.5%	20.3%	0.274	0.002	0.045	0.015
Gender	Female	60.7%	60.7%	0.001	0.004	0.023	0.008

Table 4.2 shows that the average effect for the target population, using GRF is 0.219

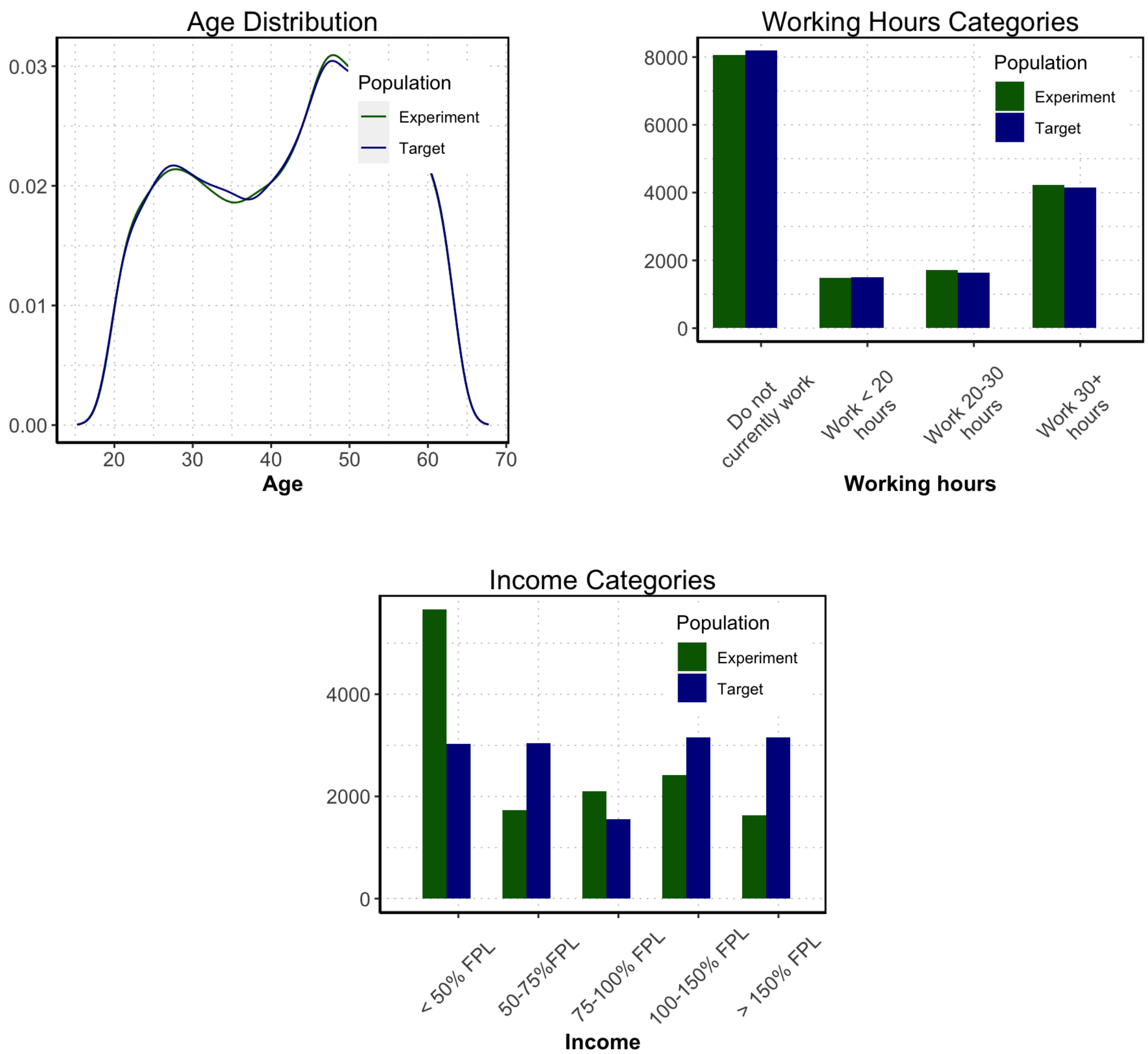

Figure (4.4) **Case 3. Distributional shift of income.**
This figure shows the distribution of age, weekly working hours, and income in the study population (green) and target population (blue). In this case the only difference between the study population and the target population is the distribution of income. The distribution of age, weekly working hours, and other variables (which are not shown here) are the same.

which is very close to the average effect in the Oregon. Moreover, the average effect estimated using LR, GBM, and RF are 0.216, 0.218, and 0.211, respectively, which confirm the GRF results.

Case 4: Changing Age, working hours, and income distributions. These three covariates are the three most important covariates in generating heterogeneity in the effect of Medicaid expansion program in Oregon. In previous cases, I demonstrated the impact of each covariate on the average effect separately. In this case, I combine all these three cases to see how the effect changes. The distributions of these covariates is shown in Figure 4.5.

Table 4.5, columns 3 and 4, show the average of age in two populations and the proportion of individuals in each category of working hours and income. The ASMD before weighting column also shows that all numbers are above 0.1 and confirms that the two populations are very different based on these three important variables. The propensity scores, estimated by all three methods, adjust for the differences and make the two populations balanced.

Table (4.5) **Case 4. Covariate means in study and target populations and Absolute Standardized Mean Difference (ASMD) before and after weighting.** This table shows the distribution of covariates in the study and target populations based on different levels of covariates. It also shows the (ASMD) of covariates before weighting and after weighting by propensity scores.

Variable	Level	Study population	Target population	ASMD (before weighting)	ASMD (after weighting)		
					Logistic	Random Forest	GBM
Age		43	32.8	0.992	0.097	0.027	0.015
Employment	Do not currently work	52.1%	30.9%	0.43	0.006	0.053	0.081
	Work < 20 hours per week	9.6%	19.8%	0.292	0.006	0.088	0.074
	Work 20-29 hours per week	11.1%	24.3%	0.355	0.001	0.069	0.090
	Work 30+ hours per week	27.2%	24.8%	0.056	0.001	0.086	0.064
Income (% FPL)	<50%	36.5%	19.5%	0.386	0.024	0.083	0.64
	50-75%	11.1%	19.6%	0.236	0.015	0.047	0.055
	75-100%	13.5%	10.0%	0.111	0.008	0.062	0.079
	100-150%	15.6%	20.4%	0.125	0.016	0.025	0.019
	Above 150%	10.5%	20.3%	0.274	0.031	0.011	0.078
Gender	Female	60.7%	60.7%	0.001	0.039	0.056	0.066

Table 4.2 shows that the average effect in the target population for this case, estimated by GRF, is 0.174. The estimated average effect with GBM is 0.177 which is the closest among the weighting methods to the estimated effect by GRF. The LR and RF also estimated the

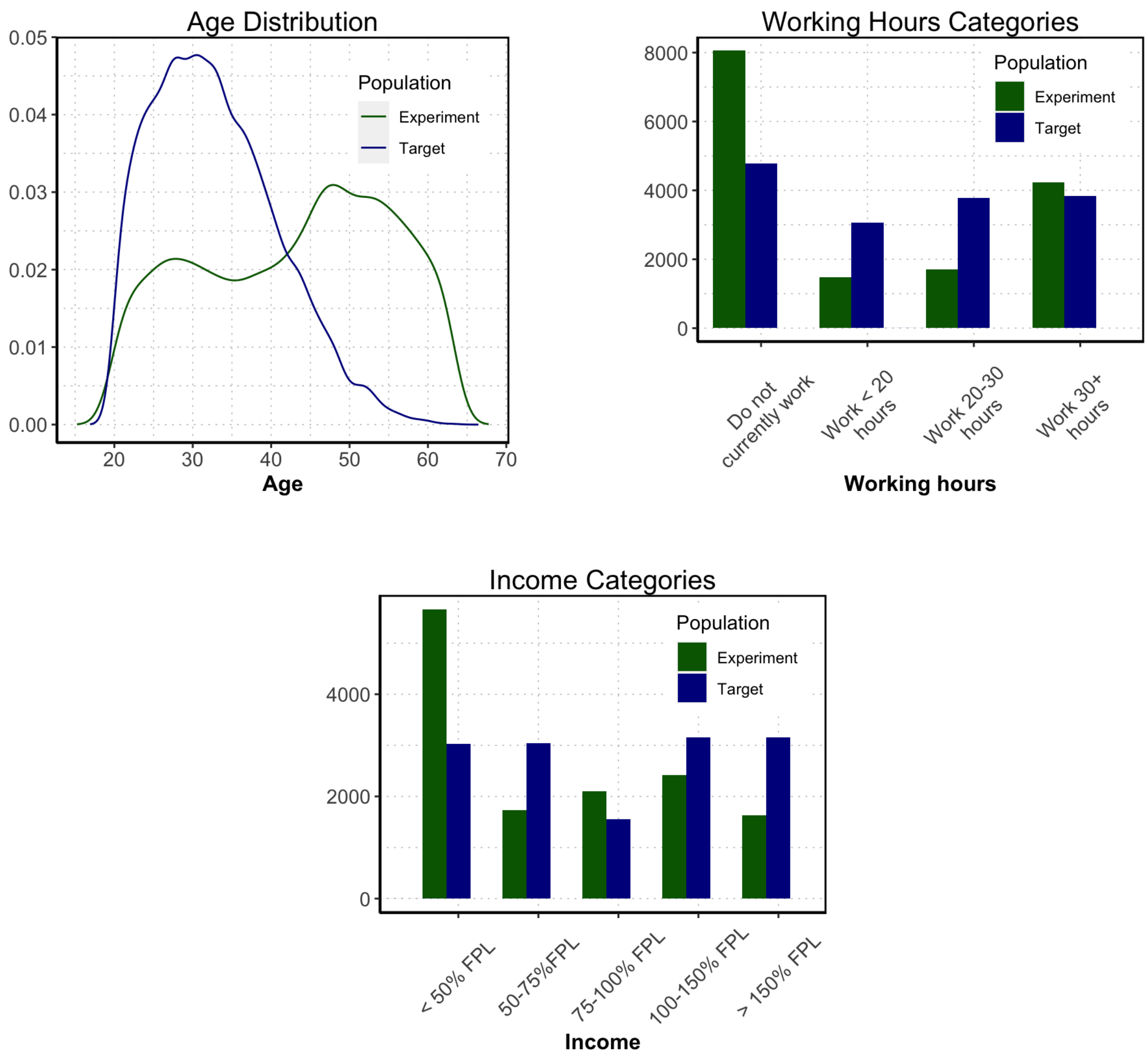

Figure (4.5) **Case 4. Distributional shift of age, working hours, and income.** This figure shows the distribution of age, weekly working hours, and income in the study population (green) and target population (blue). In this case the distribution of age, weekly working hours, and income are different between the study population and the target population. The distribution of other variables (which are not shown here) are the same.

average effect as 0.181, and 0.197 in the target population, respectively.

Case 5: Changing only age and working hours distributions. In this case, I change the distributions of age and weekly working hours of individuals in the target population such that in the population the majority of people are working more than 30 hours per week and most of them are older than the study population. I keep the distributions of other variables the same as Oregon. The distributions of age, working hours, and income variables are shown in Figure 4.6.

Table 4.6 shows that in the target population 54% of individuals are working more than 30 hours per week while in the study population about 27% of people are working more than 30 hours per week. The other three categories of working hours are different between two population. Also, this table shows that the target population on average are older than study population. The ASMD before weighting column confirms that the two populations are different based on age and all categories of working hours as their ASMDs are very large, but after weighting the ASMDs become very low (for all three methods of propensity score estimations) and show that the two populations are balanced.

Table (4.6) **Case 5. Covariate means in study and target populations and Absolute Standardized Mean Difference (ASMD) before and after weighting.** This table shows the distribution of covariates in the study and target populations based on different levels of covariates. It also shows the (ASMD) of covariates before weighting and after weighting by propensity scores.

Variable	Level	Study population	Target population	ASMD (before weighting)	ASMD (after weighting)		
					Logistic	Random Forest	GBM
Age		43	48.2	0.504	0.009	0.037	0.061
Employment	Do not currently work	52.1%	15.1%	0.848	0.005	0.015	0.076
	Work < 20 hours per week	9.6%	15.7%	0.186	0.001	0.023	0.023
	Work 20-29 hours per week	11.1%	15.0%	0.119	0.001	0.008	0.047
	Work 30+ hours per week	27.2%	54.0%	0.566	0.003	0.011	0.075
Income (% FPL)	<50%	36.5%	36.0%	0.022	0.061	0.055	0.62
	50-75%	11.1%	11.2%	0.004	0.009	0.012	0.011
	75-100%	13.5%	13.6%	0.014	0.012	0.016	0.023
	100-150%	15.6%	15.9%	0.011	0.007	0.027	0.031
	Above 150%	10.5%	10.4%	0.004	0.002	0.028	0.063
Gender	Female	60.7%	60.7%	0.001	0.013	0.010	0.019

Given these distributions, the average effects in the target population are reported in Table 4.2. GRF predicts that the average effect in the target population is 0.259 which

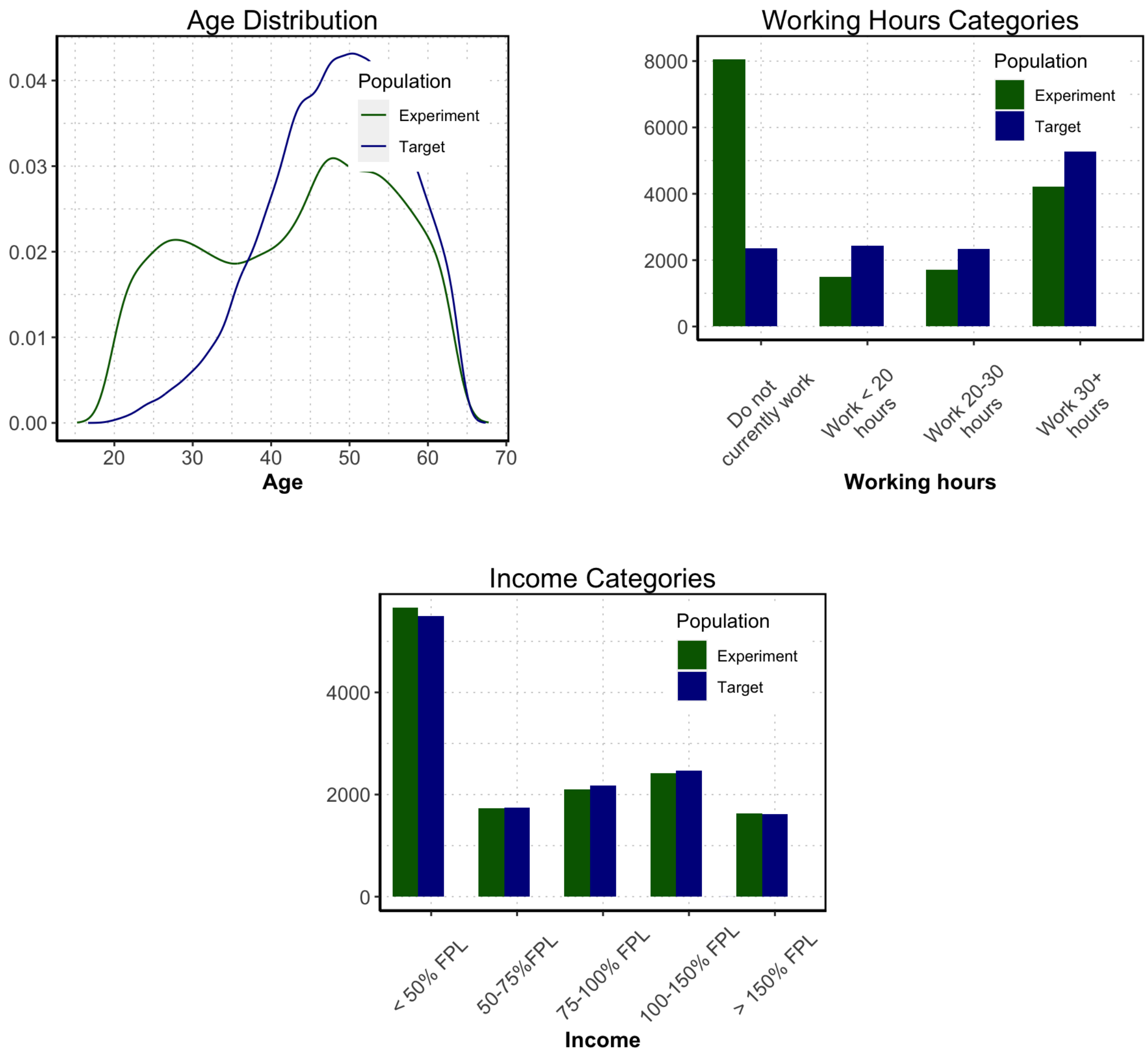

Figure (4.6) **Case 5. Distributional shift of age and weekly working hours.** This figure shows the distribution of age, weekly working hours, and income in the study population (green) and target population (blue). In this case the difference between the study population and the target population is the distributions of age and weekly working hours. In the target population a majority of people are old (around 50 years old) and most of them work more than 30 hours per week. The distribution of income, and other variables (which are not shown here) are the same.

is larger than that of Oregon experiment. This result is expected considering the Oregon experiment results. Since the effect of Medicaid enrollment on happiness in Oregon is larger for older people who work more than 30 hours per week compared to other people, the target population is expected to have a greater effect compared to Oregon experiment as this population is composed of older people who work more than 30 hours per week. The average effect from weighting methods are consistent with this result and the GBM result is the closest to the GRF result.

Case 6: Changing only the distribution of one of the less important variables. In this case, I keep the distributions of all covariates the same between the two populations except one of the less important covariates, namely the gender covariate. Distributions are shown in Figure 4.7.

Table 4.7 shows that the average of all variables are the same between the study population and the target population except the proportion of females. In the target population only 30% of population are female unlike the study population where 60% of people are female.

Table (4.7) **Case 6. Covariate means in study and target populations and Absolute Standardized Mean Difference (ASMD) before and after weighting.** This table shows the distribution of covariates in the study and target populations based on different levels of covariates. It also shows the (ASMD) of covariates before weighting and after weighting by propensity scores.

Variable	Level	Study population	Target population	ASMD	ASMD (after weighting)		
				(before weighting)	Logistic	Random Forest	GBM
Age		43	42.9	0.006	0.004	0.014	0.020
Employment	Do not currently work	52.1%	52.1%	0.017	0.008	0.006	0.013
	Work < 20 hours per week	9.6%	9.7%	0.017	0.017	0.012	0.013
	Work 20-29 hours per week	11.1%	11.1%	0.015	0.016	0.007	0.011
	Work 30+ hours per week	27.2%	26.8%	0.009	0.012	0.006	0.002
Income (% FPL)	<50%	36.5%	36.5%	0.022	0.023	0.007	0.012
	50-75%	11.1%	11.2%	0.004	0.012	0.004	0.019
	75-100%	13.5%	13.6%	0.014	0.008	0.010	0.015
	100-150%	15.6%	15.9%	0.011	0.012	0.006	0.016
	Above 150%	10.5%	10.4%	0.004	0.006	0.001	0.011
Gender	Female	60.7%	29.9%	0.649	0.003	0.010	0.032

The average effects for this scenario are shown in Table 4.2. GRF predicts the average effect for the target population very close to that of Oregon. The weighting methods also

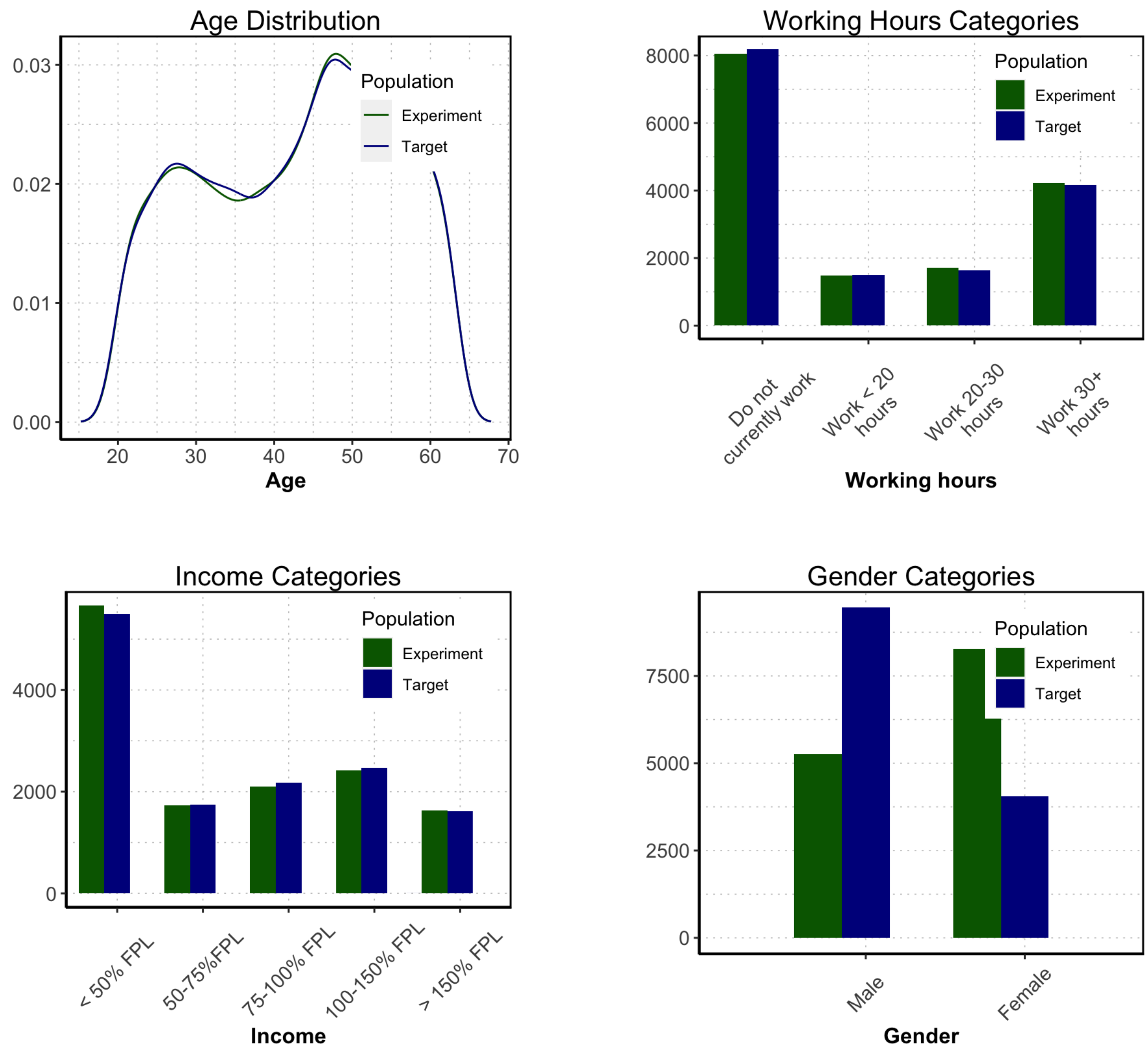

Figure (4.7) **Case 6. Distributional shift of gender.**
This figure shows the distributions of age, weekly working hours, income, and gender in the study population (green) and the target population (blue). In this case the only difference between the study population and the target population is the distribution of gender. In the target population a majority of people are male. The distributions of age, weekly working hours, income, and other variables (which are not shown here) are the same.

show the same results. This confirms that the gender of the participants does not play a big role in defining the effect of Medicaid expansion program on the happiness of the participant.

Case 7: Changing the correlation between covariates, and keeping covariate means and standard deviation the same. In previous cases (case 1 to case 6), the correlation matrix for covariates in the target population is the same as the correlation matrix for covariates in the study population, but the mean and standard deviation for some covariates are different. In this case, I keep the mean and standard deviation of all the covariates the same, but the correlation matrix of covariates are different in two populations. Figure 4.8 on the left panel shows the correlation in the study population and the right panel shows the correlation matrix in the target population.

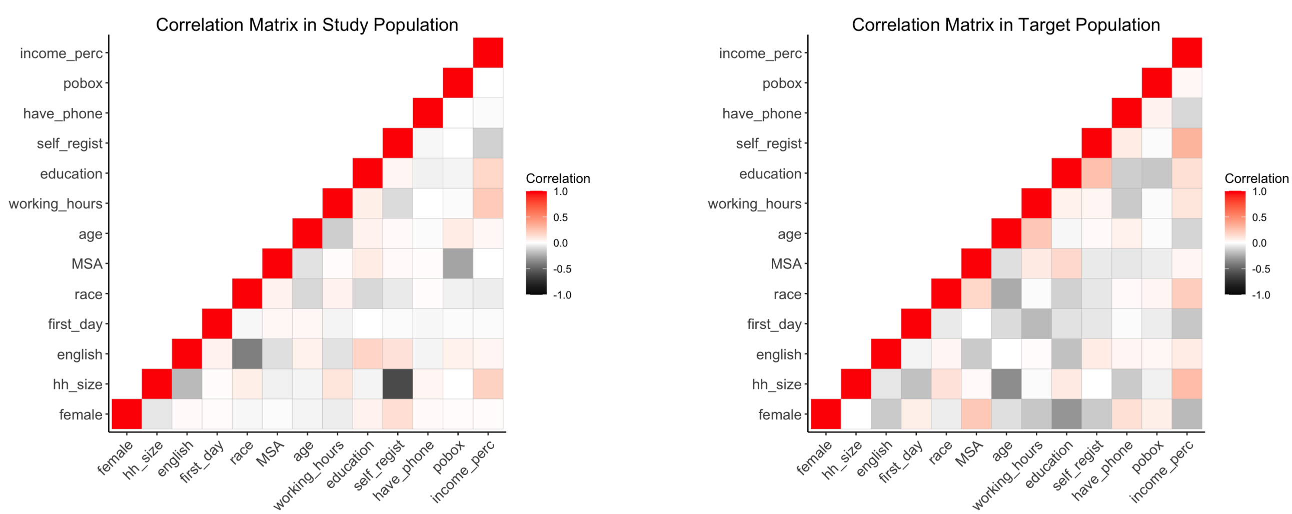

Figure (4.8) **Case 7. Changing the Correlation Matrix.**
This figure shows the correlation matrix in the study population (left panel) and target population (right panel). Covariates in the study population are more correlated compared to target population.

Using this new correlation matrix I generate the covariates in the target population such that they are correlated based on the defined correlation matrix. The covariates in the target population have the same mean and standard deviation as covariates in the study population. The estimated average effect in the target population is reported in Table 4.2.

GRF predicts that the average effect in the target population is 0.218 and the average effects from the weighting methods are all the same and is 0.214. These numbers show that changing the correlation between covariates does not affect the average effect in the target population.

Chapter 5

DISCUSSION AND CONCLUSION

This dissertation aimed to provide new evidence of the effect of Medicaid expansion program on several outcome variables beyond physical health and generalize the estimated effects to a population where the Medicaid expansion program is yet to be implemented.

In Chapter 2, I explained the details about the Oregon Health Insurance Experiment setting and the dataset that I used. Moreover, I described the challenges in estimating the causal effect of Medicaid and introduced the newly developed causal Machine Learning method, Generalized Random Forest (GRF), by Athey and Wager (2019). I pointed out the traditional econometrics methods for estimating the heterogeneous treatment effect and the advantages of GRF over these methods. In Chapter 3, I presented the estimated heterogeneous effects of Medicaid expansion program on several outcome variables, namely self-reported happiness, self-reported out-of-pocket costs, and self-reported depression using GRF. In addition to that I provided some evidence that shed light on the potential mechanisms through which Medicaid may affect these outcomes. In Chapter 4, I used the estimated effects from Chapter 3, the heterogeneous effects of Medicaid expansion program in Oregon on happiness, to estimate the average effect of Medicaid on happiness for people living in another population where the Medicaid expansion program has not implemented yet. The goal was to show what would be the average effect in another population with different distributions of covariates, if this population are given access to Medicaid expansion program in the same way that was given to Oregonians. Due to lack of real data for the target population I simulated this population using a Copula approach. I generated the data in the target population such that the correlation matrix between covariates is the same as the correlation matrix in Oregon while the mean of the variables are different. Moreover, I kept the marginal distributions of covariates the same and changed the correlation between them. For each scenario I estimated the average effect of Medicaid in the target population.

5.1 Concluding Remarks

Estimating the effect of Medicaid is challenging because it is correlated with unobservable variables. The Oregon's 2008 lottery-allocated access to Medicaid provided an opportunity to study the causal impact of Medicaid. In the literature almost all studies on Medicaid estimate the Average effect using two-stage least square method. The drawback with this method is that it hides the differences between individuals and the way they are affected by Medicaid. Moreover, estimating the average effect raises the questions about the mechanisms through which Medicaid may affect outcomes. My contribution by this study is to address these issues and estimate the Medicaid's effect as a function of individuals characteristics and provide heterogeneous effect of Medicaid on self reported happiness, self reported out-of-pocket medical cost, and self-reported depression.

In this study, using Generalized Random Forest (GRF) model, I found that age, weekly working hours, and urbanicity create substantial heterogeneity in Medicaid's effect. Medicaid causes older adults who work more that 30 hours per week to be happier and the possible driving factor of this effect is the pen-up health care demand. Moreover, Medicaid decreases the out-of-pocket medical costs of rural population by a greater amount and this is likely due to the differential health care competition in urban and rural areas before implementing Medicaid. In addition to that Medicaid is very successful in decreasing the depression of younger adults and the plausible factor that can explain this effect is the education-related efficient use of health services among young adults.

The goal of the policy makers by conducting various experiments is to use the results from the experiments to find out whether they have to adopt the policy that they have examined in the experiment in another setting or not. In practice, the experimental findings are not generalizable or transportable to a new population due to differences between individuals in two populations. In this study, I used several methods to transport the results from the Oregon experiment to a distinct target population which has no overlap with individuals in Oregon experiment. This population could be any other state that has not implemented the Medicaid expansion program yet. Since the individual-level data for the target population was not available, I simulated the target population with synthetic distributions using a

Copula approach.

I employed two methods for transporting the effects from the Oregon experiment to the target population. I used trained GRF model on the Oregon experiment's data and inverse propensity score weighting. To estimate the propensity scores I adopted Logistic Regression, Random Forest, and Gradient Boosting models. Surprisingly when I estimate propensity scores by logistic regression the results were not very different from the random forest and generalized boosted model. The results show that linear models do not perform as poorly as one might expect. However, without having the real data on target population I cannot assess the various estimators of the average effect in the target population. I considered different scenarios of covariate distributions in the target population and estimated the average effect of Medicaid. Intuitively the results makes sense. If the individual-level data for states that have not implemented the Medicaid expansion program are available I could compare my results with the real data as a benchmark.

In this study I showed that in transporting results it is very important to control for covariates that have important role in effect heterogeneity as they can change the effect significantly.

5.2 Limitations

I acknowledge the following limitations in my study. First, in estimating the heterogeneous effect of Medicaid, the outcome variables are selected from the follow-up survey. This survey was conducted only one year after the lottery drawings. As such, The long-term effects of Medicaid may differ from my findings in this study. Second, there are other mechanisms that can drive the effect of Medicaid which I have not considered in this study. Third, in transporting results I relied on a strong assumption that the compliance rate is the same between two populations which might not be true in the real world. Fourth, I presented the average effects in the target population based on simulated data. The results would differ with the real data.

Despite these limitations, this study provides new perspective on Medicaid's effect. My results help policymakers of the States, in which the Medicaid expansion program has not yet implemented, to better understand the effect of Medicaid expansion and decide if they should expand this program. It is also helpful for policymakers to design or redesign a targeted Medicaid program.

Bibliography

Allen, H., Baicker, K., Taubman, S., Wright, B., and Finkelstein, A. (2013). "The oregon health insurance experiment: when limited policy resources provide research opportunities." *Journal of health politics, policy and law*, 38(6), 1183–1192.

Angrist, J. D. and Imbens, G. W. (1995). "Identification and estimation of local average treatment effects." *Report no.*, National Bureau of Economic Research.

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). "Identification of causal effects using instrumental variables." *Journal of the American statistical Association*, 91(434), 444–455.

Athey, S. and Imbens, G. (2016). "Recursive partitioning for heterogeneous causal effects." *Proceedings of the National Academy of Sciences*, 113(27), 7353–7360.

Athey, S., Tibshirani, J., Wager, S., et al. (2019). "Generalized random forests." *The Annals of Statistics*, 47(2), 1148–1178.

Athey, S. and Wager, S. (2019). "Estimating treatment effects with causal forests: An application." *Observational Studies*, 5(2), 37–51.

Baicker, K., Allen, H. L., Wright, B. J., and Finkelstein, A. N. (2017). "The effect of medicaid on medication use among poor adults: evidence from oregon." *Health Affairs*, 36(12), 2110–2114.

Baicker, K., Allen, H. L., Wright, B. J., Taubman, S. L., and Finkelstein, A. N. (2018). "The effect of medicaid on management of depression: evidence from the oregon health insurance experiment." *The Milbank Quarterly*, 96(1), 29–56.

Baicker, K., Finkelstein, A., Song, J., and 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), 322–28.

Baicker, K., Taubman, S. L., Allen, H. L., Bernstein, M., Gruber, J. H., Newhouse, J. P., Schneider, E. C., Wright, B. J., Zaslavsky, A. M., and Finkelstein, A. N. (2013). “The oregon experiment—effects of medicaid on clinical outcomes.” *New England Journal of Medicine*, 368(18), 1713–1722.

Breiman, L. (2001). “Random forests.” *Machine learning*, 45(1), 5–32.

Buchanan, A. L., Hudgens, M. G., Cole, S. R., Mollan, K. R., Sax, P. E., Daar, E. S., Adimora, A. A., Eron, J. J., and Mugavero, M. J. (2018). “Generalizing evidence from randomized trials using inverse probability of sampling weights.” *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 181(4), 1193–1209.

Cole, S. R. and Stuart, E. A. (2010). “Generalizing evidence from randomized clinical trials to target populations: the actg 320 trial.” *American journal of epidemiology*, 172(1), 107–115.

Degtiar, I. and Rose, S. (2021). “A review of generalizability and transportability.” *arXiv preprint arXiv:2102.11904*.

Epstein, R. J. (1989). “The fall of ols in structural estimation.” *Oxford Economic Papers*, 41(1), 94–107.

Fertig, A. R., Carlin, C. S., Ode, S. B., and Long, S. K. (2016). “Pent-up demand after the affordable care act.” *Journal of Patient-Centered Research and Reviews*, 3(3), 209.

Finkelstein, A., Taubman, S., Wright, B., Bernstein, M., Gruber, J., Newhouse, J. P., Allen, H., Baicker, K., and Group, O. H. S. (2012). “The oregon health insurance experiment: evidence from the first year.” *The Quarterly journal of economics*, 127(3), 1057–1106.

Finkelstein, A. N., Taubman, S. L., Allen, H. L., Wright, B. J., and Baicker, K. (2016). “Effect of medicaid coverage on ed use—further evidence from oregon’s experiment.” *N Engl J Med*, 375(16), 1505–1507.

Flay, B. R., Biglan, A., Boruch, R. F., Castro, F. G., Gottfredson, D., Kellam, S., Mościcki,

E. K., Schinke, S., Valentine, J. C., and Ji, P. (2005). "Standards of evidence: Criteria for efficacy, effectiveness and dissemination." *Prevention science*, 6(3), 151–175.

Fletcher, J. M. and Frisvold, D. E. (2009). "Higher education and health investments: does more schooling affect preventive health care use?." *Journal of human capital*, 3(2), 144–176.

Greenhouse, J. B., Kaizar, E. E., Kelleher, K., Seltman, H., and Gardner, W. (2008). "Generalizing from clinical trial data: a case study. the risk of suicidality among pediatric antidepressant users." *Statistics in medicine*, 27(11), 1801–1813.

Haneuse, S., Schildcrout, J., Crane, P., Sonnen, J., Breitner, J., and Larson, E. (2009). "Adjustment for selection bias in observational studies with application to the analysis of autopsy data." *Neuroepidemiology*, 32(3), 229–239.

Holland, P. W. (1986). "Statistics and causal inference." *Journal of the American statistical Association*, 81(396), 945–960.

Horvitz, D. G. and Thompson, D. J. (1952). "A generalization of sampling without replacement from a finite universe." *Journal of the American statistical Association*, 47(260), 663–685.

Imbens, G. W. and Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Kern, H. L., Stuart, E. A., Hill, J., and Green, D. P. (2016). "Assessing methods for generalizing experimental impact estimates to target populations." *Journal of research on educational effectiveness*, 9(1), 103–127.

Kim, S. and Koh, K. (2018). "Does health insurance make people happier? evidence from massachusetts' healthcare reform.

Lan, W., Zhong, P.-S., Li, R., Wang, H., and Tsai, C.-L. (2016). "Testing a single regression coefficient in high dimensional linear models." *Journal of econometrics*, 195(1), 154–168.

Lesko, C. R., Buchanan, A. L., Westreich, D., Edwards, J. K., Hudgens, M. G., and Cole, S. R. (2017). "Generalizing study results: a potential outcomes perspective." *Epidemiology (Cambridge, Mass.)*, 28(4), 553.

Levy, H. and Meltzer, D. (2008). "The impact of health insurance on health." *Annu. Rev. Public Health*, 29, 399–409.

List, J. A., Shaikh, A. M., and Xu, Y. (2019). "Multiple hypothesis testing in experimental economics." *Experimental Economics*, 22(4), 773–793.

McGee, R. E. and Thompson, N. J. (2015). "Peer reviewed: unemployment and depression among emerging adults in 12 states, behavioral risk factor surveillance system, 2010." *Preventing chronic disease*, 12.

National Academies of Sciences, Engineering, and Medicine (2018). "Health-care utilization as a proxy in disability determination.

National Center for Health Statistics and Center For Disease Control and Preventio (2017). "Health, united states, 2016, with chartbook on long-term trends in health.

Pearl, J. and Bareinboim, E. (2011). "Transportability of causal and statistical relations: A formal approach." *Twenty-fifth AAAI conference on artificial intelligence*.

Qiu, Y., Tao, J., and Zhou, X.-H. (2021). "Inference of heterogeneous treatment effects using observational data with high-dimensional covariates." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.

Rosenbaum, P. R. and Rubin, D. B. (1983). "The central role of the propensity score in observational studies for causal effects." *Biometrika*, 70(1), 41–55.

Sekhon, J. S. (2008). "The neyman-rubin model of causal inference and estimation via matching methods." *The Oxford handbook of political methodology*, 2, 1–32.

Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton, Mifflin and Company.

Shakya, S. and McCullough, J. (2020). "Heterogeneous treatment effects of medicaid and efficient policies: Evidence from the oregon health insurance experiment." *9th Annual Conference of the American Society of Health Economists*, ASHECON.

Stock, J. H. and Trebbi, F. (2003). "Retrospectives: who invented instrumental variable regression?." *Journal of Economic Perspectives*, 17(3), 177–194.

Stuart, E. A., Cole, S. R., Bradshaw, C. P., and Leaf, P. J. (2011). "The use of propensity scores to assess the generalizability of results from randomized trials." *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 174(2), 369–386.

Stuart, E. A. and Rhodes, A. (2017). "Generalizing treatment effect estimates from sample to population: A case study in the difficulties of finding sufficient data." *Evaluation review*, 41(4), 357–388.

Taubman, S. L., Allen, H. L., Wright, B. J., Baicker, K., and Finkelstein, A. N. (2014). "Medicaid increases emergency-department use: evidence from oregon's health insurance experiment." *Science*, 343(6168), 263–268.

Tipton, E. (2013). "Improving generalizations from experiments using propensity score subclassification: Assumptions, properties, and contexts." *Journal of Educational and Behavioral Statistics*, 38(3), 239–266.

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.

Weisberg, H. I., Hayden, V. C., and Pontes, V. P. (2009). "Selection criteria and generalizability within the counterfactual framework: explaining the paradox of antidepressant-induced suicidality?." *Clinical trials*, 6(2), 109–118.

Zhang, Z., Kim, H. J., Lonjon, G., Zhu, Y., et al. (2019). "Balance diagnostics after propensity score matching." *Annals of translational medicine*, 7(1).

Appendix A

SUPPLEMENT TO CHAPTER 3

A.1 Proofs for section 3.1

$$Y_i = \nu(X_i) + \theta(X_i)D_i + \epsilon_i$$

$$E[Y_i|Z_i, X_i] = E[\nu(X_i)|Z_i, X_i] + E[\theta(X_i)D_i|Z_i, X_i] + E[\epsilon_i|Z_i, X_i]$$
$$E[Y_i|Z_i, X_i] = \nu(X_i) + \theta(X_i)E[D_i|Z_i, X_i] + 0$$

To get the true parameter we need $E[\epsilon_i|Z_i, X_i] = 0$ and we assume $Z \in 0, 1$:

$$Z = 1 \Rightarrow E[Y_i|Z_i = 1, X_i] = \nu(X_i) + \theta(X_i)E[D_i|Z_i = 1, X_i]$$
$$Z = 0 \Rightarrow E[Y_i|Z_i = 0, X_i] = \nu(X_i) + \theta(X_i)E[D_i|Z_i = 0, X_i]$$

$\theta(X_i)$ can be calculated by

$$\theta(X_i) = \frac{E[Y_i|Z_i = 1, X_i] - E[Y_i|Z_i = 0, X_i]}{E[D_i|Z_i = 1, X_i] - E[D_i|Z_i = 0, X_i]} \quad (A.1)$$

The numerator in equation A.1 can be writted as

$$\begin{aligned} E[Y_i|Z_i = 1, X_i] - E[Y_i|Z_i = 0, X_i] =& E[(Y_{1i} - Y_{0i})(D_{1i} - D_{0i})|X_i] \\ =& E[(Y_{1i} - Y_{0i}) \times 1|D_{1i} - D_{0i} = 1, X_i] \times Pr(D_{1i} - D_{0i} = 1|X_i) \\ +& E[(Y_{1i} - Y_{0i}) \times 0|D_{1i} - D_{0i} = 0, X_i] \times Pr(D_{1i} - D_{0i} = 0|X_i) \\ +& E[(Y_{1i} - Y_{0i}) \times -1|D_{1i} - D_{0i} = -1, X_i] \times Pr(D_{1i} - D_{0i} = -1|X_i) \end{aligned}$$

By monotonicity assumption the numerator is equal to

$$E[Y_i|Z_i = 1, X_i] - E[Y_i|Z_i = 0, X_i] = E[Y_{1i} - Y_{0i}|D_{1i} - D_{0i} > 0, X_i] \times Pr(D_{1i} - D_{0i} > 0|X_i)$$

The denominator in equation A.1 can be written as

$$E[D_i|Z_i = 1, X_i] - E[D_i|Z_i = 0, X_i]$$
$$= E[D_{0i} + (D_{1i} - D_{0i})Z_i|Z_i = 1, X_i] - E[D_{0i} + (D_{1i} - D_{0i})Z_i|Z_i = 0, X_i]$$
$$= E[D_{1i}|X_i] - E[D_{0i}|X_i] = E[D_{1i} - D_{0i}|X_i]$$
$$= 1 \times Pr(D_{1i} - D_{0i} = 1|X_i) + 0 \times Pr(D_{1i} - D_{0i} = 0|X_i) + -1 \times Pr(D_{1i} - D_{0i} = -1|X_i)$$

By monotonicity assumption this equation is equal to

$$= Pr(D_{1i} - D_{0i} > 0|X_i)$$

Therefore

$$\theta(X_i) = \frac{E[Y_i|Z_i = 1, X_i] - E[Y_i|Z_i = 0, X_i]}{E[D_i|Z_i = 1, X_i] - E[D_i|Z_i = 0, X_i]} = E[Y_{1i} - Y_{0i}|D_{1i} - D_{0i} > 0, X_i]$$

A.2 Self-reported Happiness

In Chapter 3, I showed the heterogeneity in effects of Medicaid on happiness in multiple dimensions. In this section I add another dimension to the heterogeneity in effects of Medicaid on happiness and show the heterogeneity with respect to age, working hours (represented by colors), and income category (represented in different plots). As Figure A.1 shows, there are five categories for income and each subplot shows the effects for each income category.

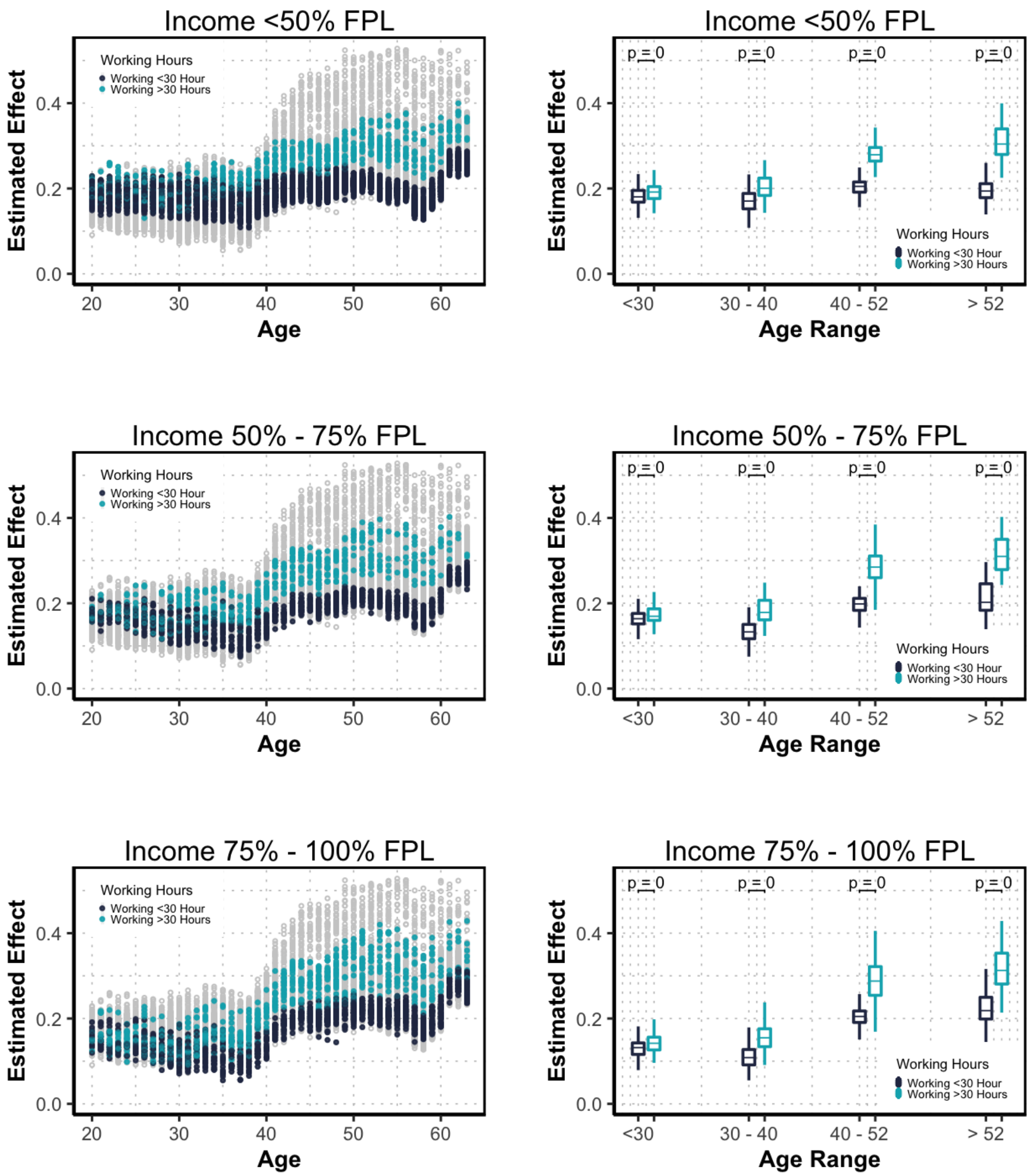

Figure (A.1) **Heterogeneity in Medicaid's Effects on Happiness, by age, working hours and income.**

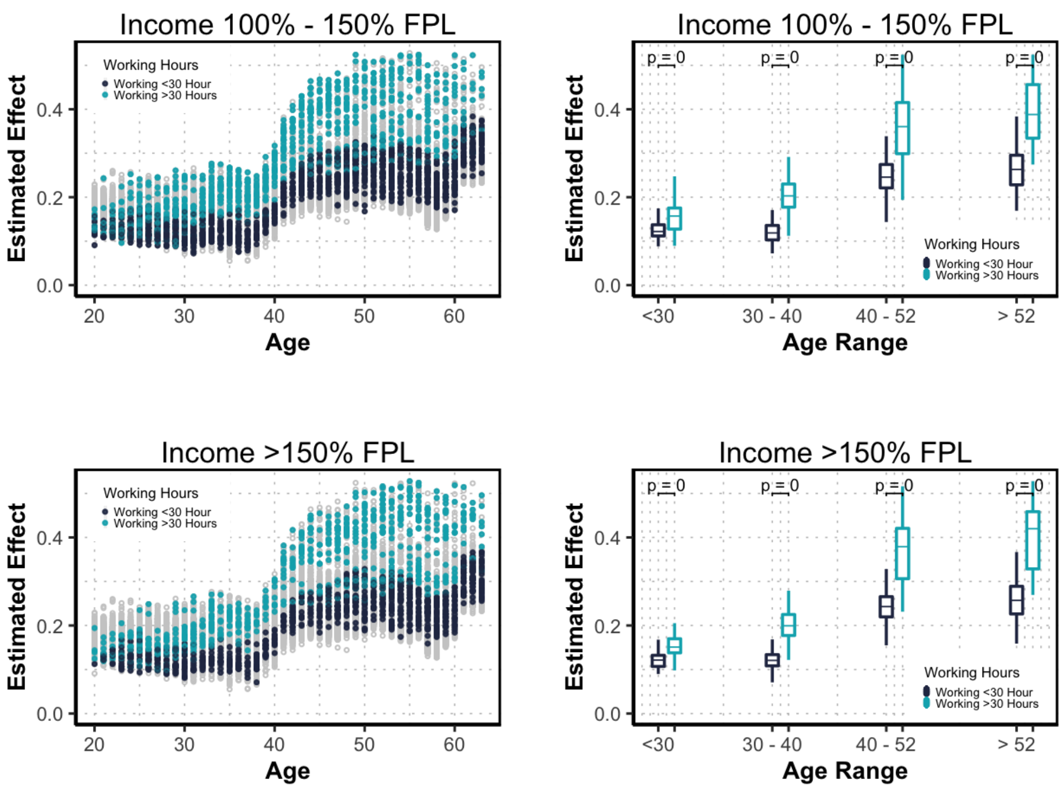

Figure (A.1) (cont.)

A.3 Self-reported Out-of-Pocket Medical Costs

In Chapter 3, I showed the heterogeneity in effects of Medicaid on out-of-pocket medical costs in multiple dimensions. In this section I include more details from GRF for this outcome variable. Moreover, I add another dimension to the heterogeneity in effects of Medicaid on out-of-pocket medical costs and show the heterogeneity with respect to age, MSA (represented by colors), and income category (represented in different plots). As Figure A.6 shows, there are five categories for income and each subplot shows the effects for each income category.

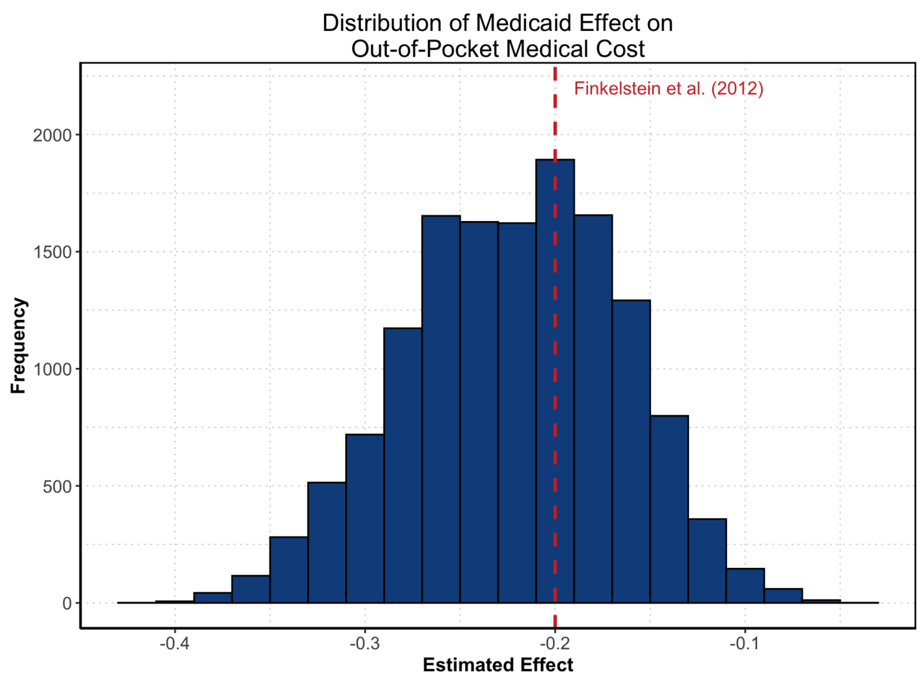

Figure (A.2) **Histogram of estimated Medicaid's effect on OOP Costs by GRF.** This plot shows the distribution of individual-level estimated effect of Medicaid on out-of-pocket costs. The effect for all individuals is negative and varies between -0.43 and -0.15. The mean of this distribution is -0.22. The red dashed line shows the estimated average effect of Medicaid on out-of-pocket costs reported in Finkelstein et al. (2012).

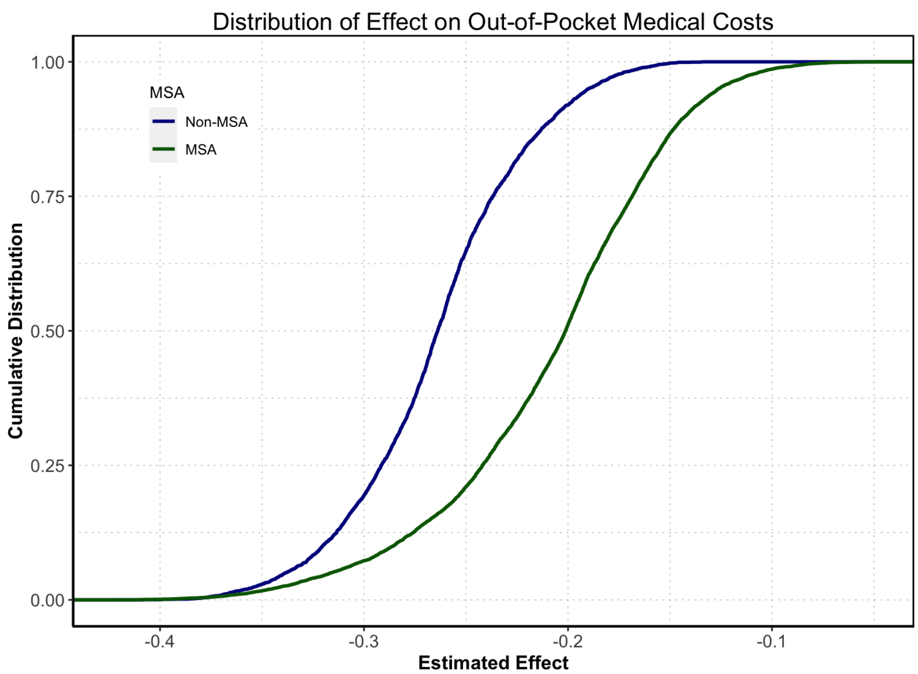

Figure (A.3) **CDF of Medicaid's Effect on Out-of-pocket Costs for MSA.** This figure shows the distribution of effect for MSA and non-MSA.

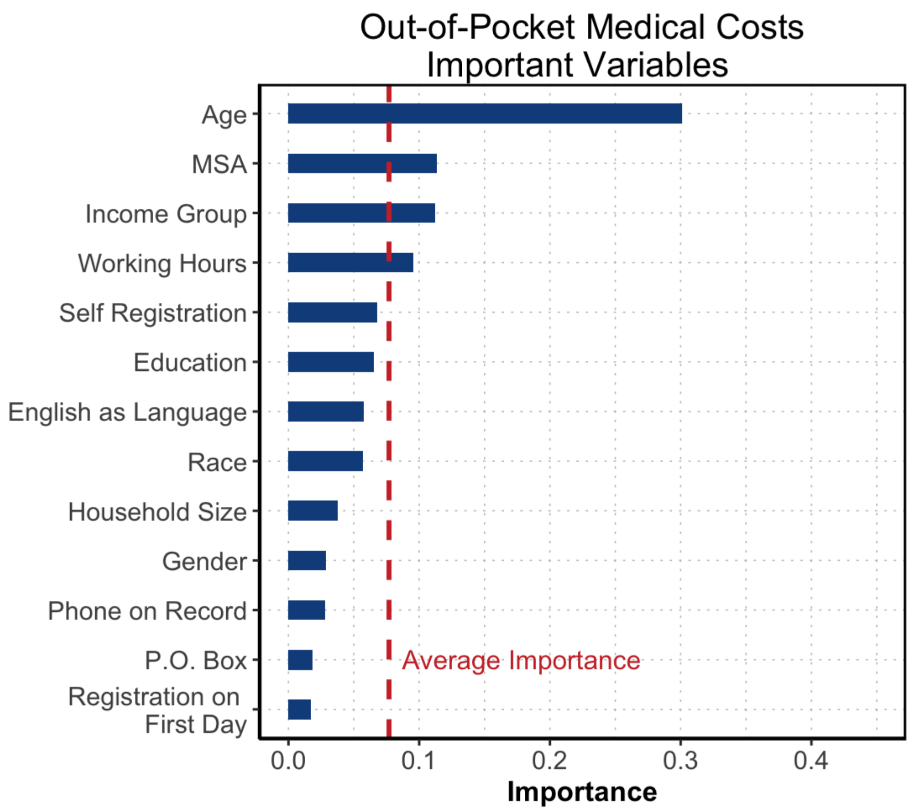

Figure (A.4) **The Importance Matrix for OOP Costs from the GRF.**
This figure represents the importance measure of each variable in the GRF algorithm in estimating the effect of Medicaid on out-of-pocket costs. It determines how often the GRF algorithm chooses a variable for splitting the covariate space across all trees in the forest.

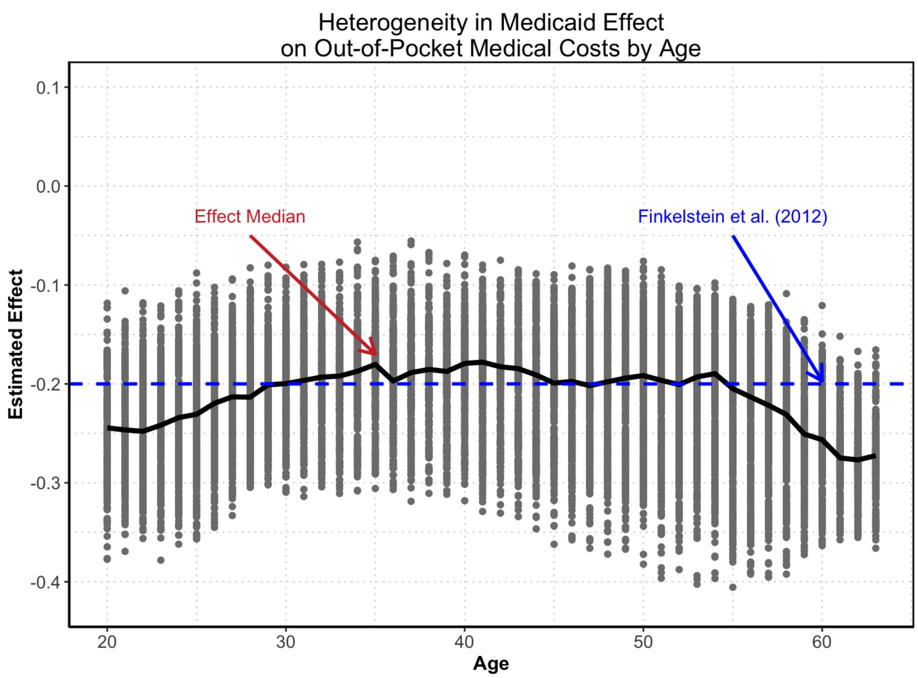

Figure (A.5) **Distribution of Medicaid's effect on OOP Costs Versus Age.** This figure represents how the estimated effect of Medicaid on out-of-pocket costs varies with respect to age. To have a better sense of effect's variability, I added the effect's median line (the solid black line) across the age range to this figure. The Finkelstein et al. (2012)'s result is shown with the blue dashed line.

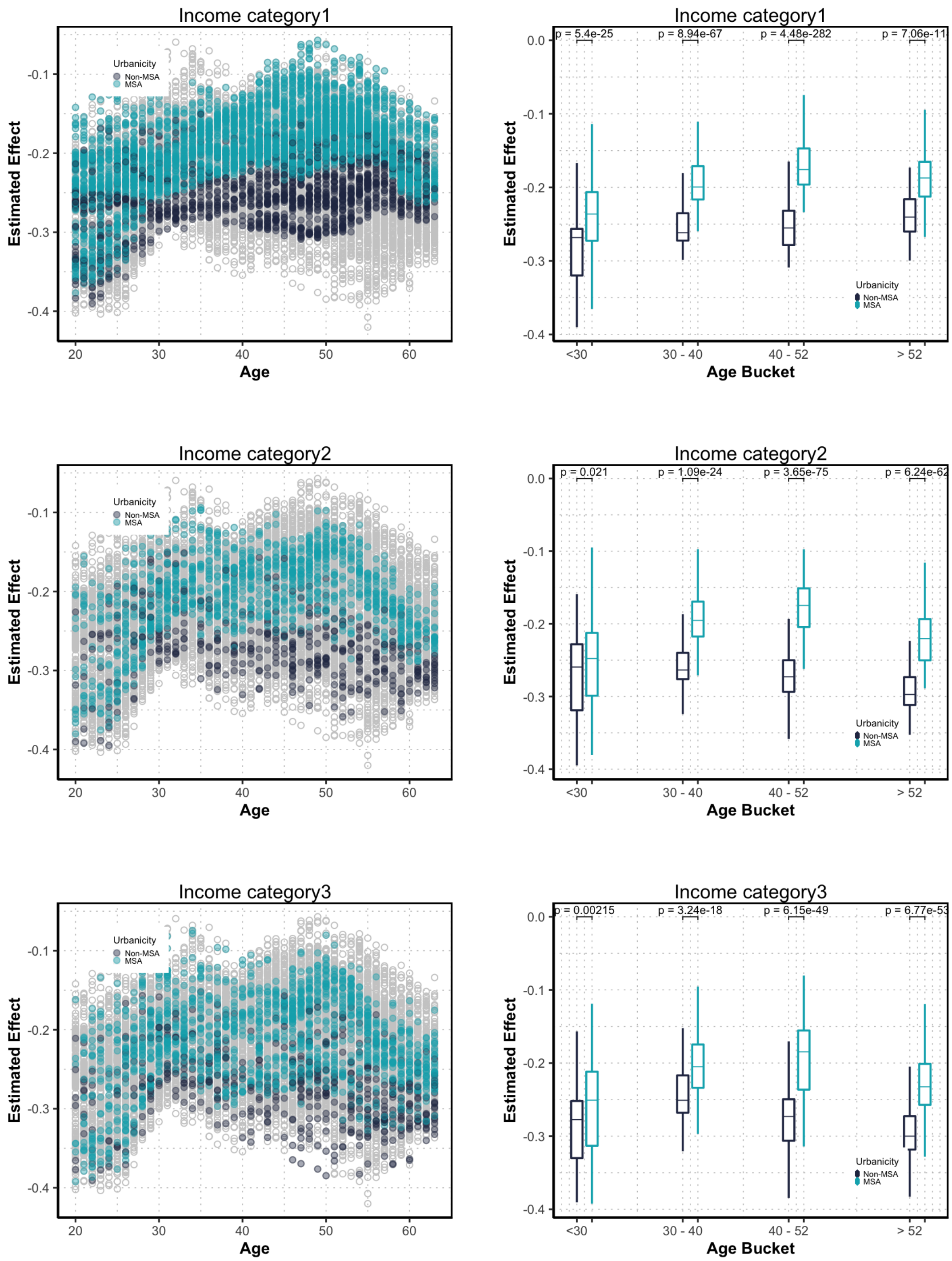

Figure (A.6) **Heterogeneity in Medicaid's Effects on OOP Costs, by Age, Working Hours and Income.**

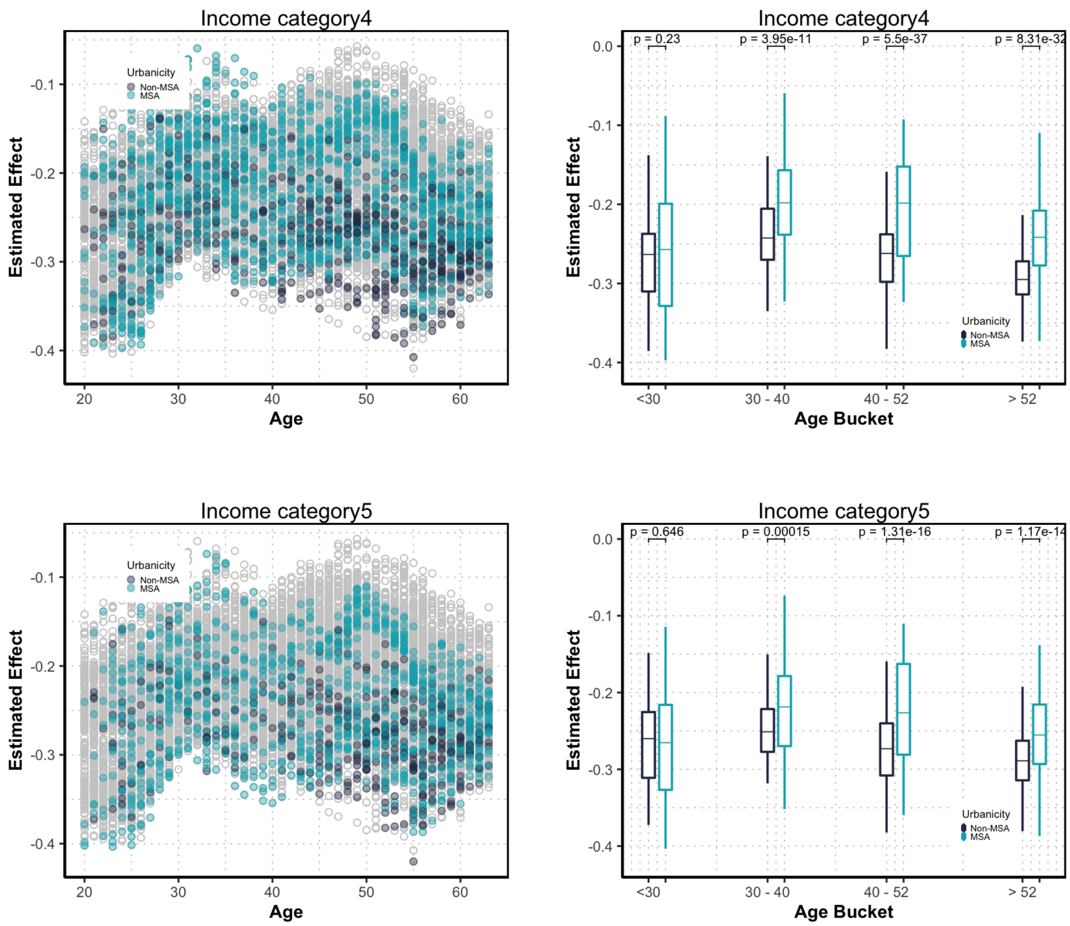

Figure (A.6) (cont.)

A.4 Self-reported Depression

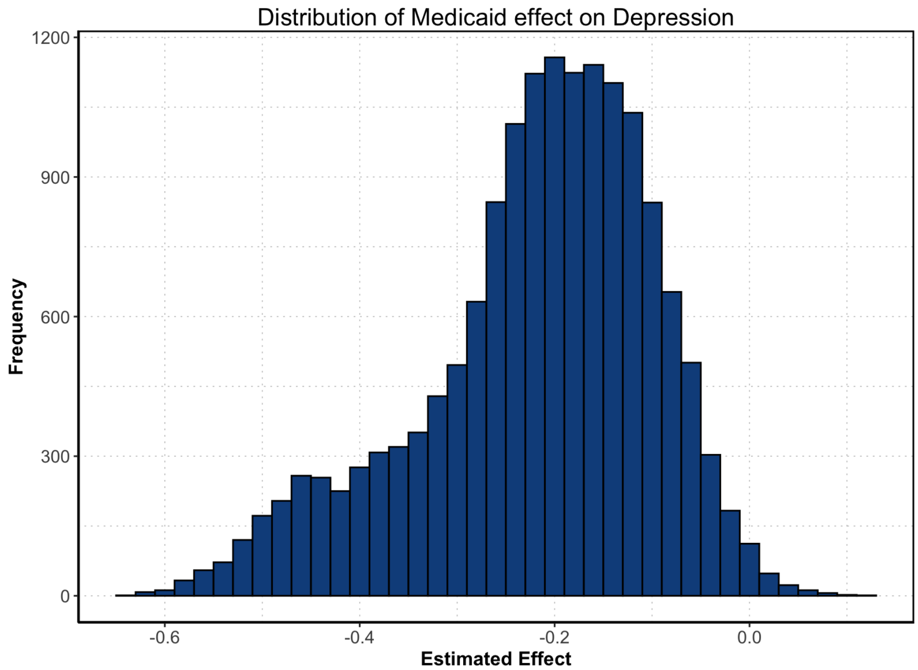

Figure (A.7) **Histogram of estimated Medicaid's effect on Depression by GRF.** This plot shows the distribution of individual-level estimated effect of Medicaid on Depression. The effect is mostly negative and varies between -0.63 and 0.1. The mean of this distribution is -0.21.

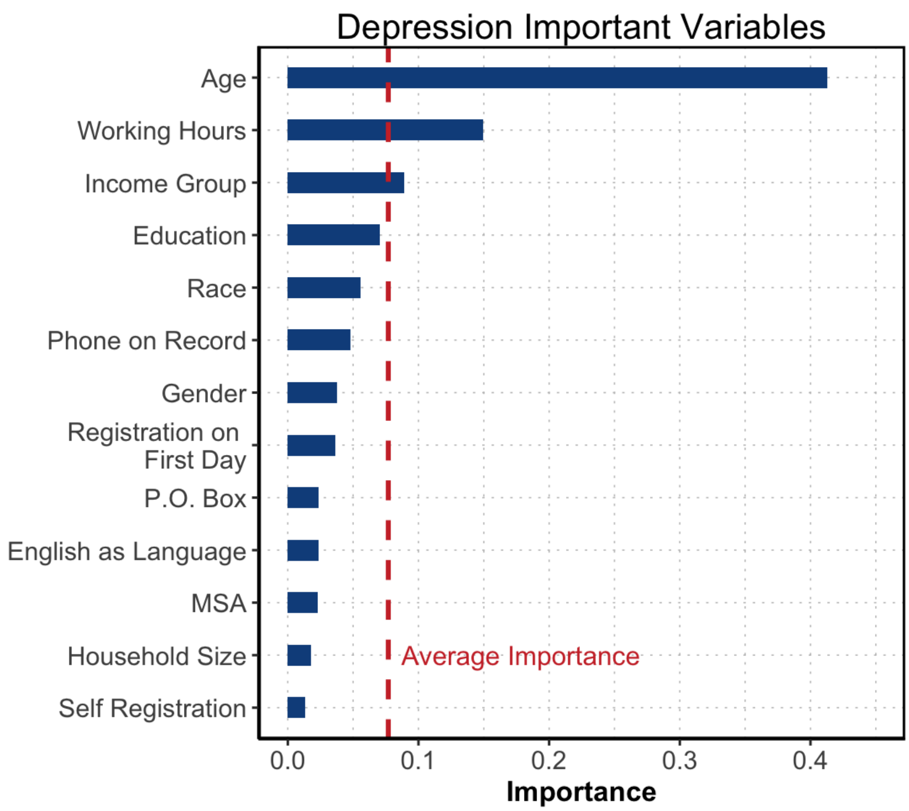

Figure (A.8) **The Importance Matrix for Depression from the GRF.**
This figure represents the importance measure of each variable in the GRF algorithm in estimating the effect of Medicaid on Depression. It determines how often the GRF algorithm chooses a variable for splitting the covariate space across all trees in the forest.

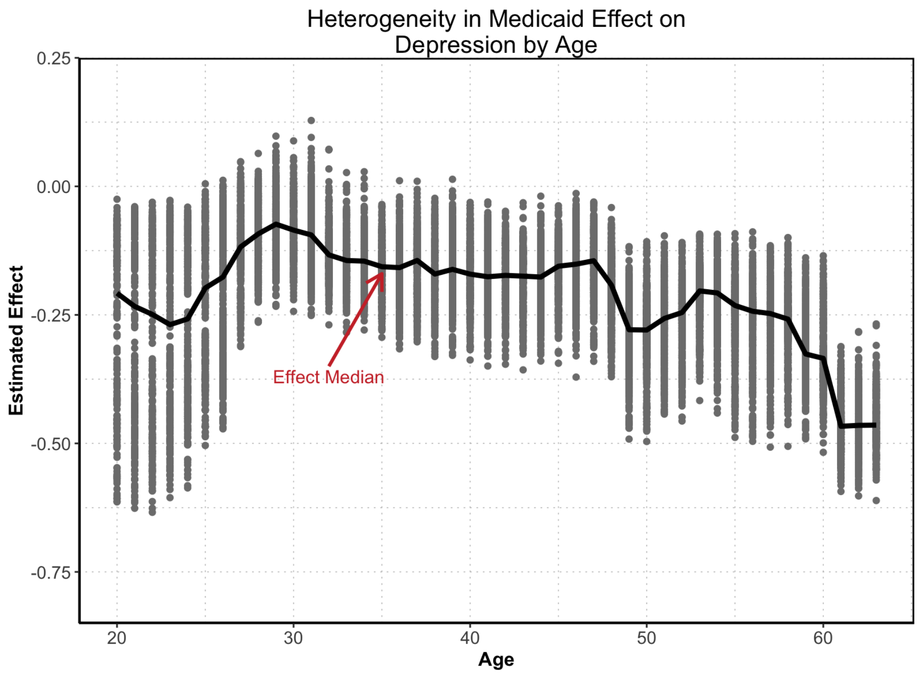

Figure (A.9) **Distribution of Medicaid's effect on Depression Versus Age.** This figure represents how the estimated effect of Medicaid on Depression varies with respect to age. To have a better sense of effect's variability, I added the effect's median line (the solid black line) across the age range to this figure.