UNIVERSITÀ DEGLI STUDI DI MILANO

FACOLTÀ DI SCIENZE POLITICHE,
ECONOMICHE E SOCIALI

Corso di Laurea Magistrale in Data Science and Economics

The effects of the *Avancemos* Conditional Cash Transfer Program in Costa Rica, evaluation using Machine Learning

Relatore:
Prof.
Silvia SALINI

Candidato:
Ricardo MURILLO
RAPSO
Mat. 923159

Correlatore:
Prof.
Giancarlo MANZI

ANNO ACCADEMICO 2021/2022

Contents

List of Figures

List of Tables

Abstract

The objective of this project is to use a series of Machine Learning Methods as a powerful approach for solving the problem of policy evaluation of the conditional money transfer policy *Avancemos* in Costa Rica, studied for the years 2010 to 2021, and using as a base the observational data *National Household Survey* for these years. The variable of interest, from which the impact of the policy will be studied is child labor, while also controlling for the endogeneity of the program participation, and the preservation of the a great amount of variables. For the estimation of the treatment effect, the Double Machine Learning Method is used, with Lasso, Adaboost, Random Forest and Tree Models. The effect of the program is mildly positive with a positive effect of around 1% to 4% of moving reducing the possibility of a child 12-17 years old, to choose working over school.

Keywords:

Machine Learning, Causality, Policy Evaluation, Double Machine Learning.

Declaration

I declare that this material, which will be now submitted for assessment, is entirely my own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my work. I understand that plagiarism,collusion, and copying are grave and serious offences in the university and accept the penalties that would be imposed should I engage in plagiarism, collusion or copying. This assignment, or any part of it, has not been previously submitted by me or any other person for assessment on this or any other course of study.

Chapter 1

Introduction

Public policy is typically designed to benefit the outcomes of the most vulnerable population, and it has goals aimed at improving the conditions of persons on topics such as healthcare, education, self-awareness, employment, transportation and others. Determining whether policies are effective on reaching their goals, and the improvements that can be reached is crucial for public administrators and the citizenship in general. Impact evaluation can make assessments on evidence based policy, giving data supported answers on the impact of policies [3].

One of the most interesting questions in research has always been about relationships of cause and effect, which is the main topic of impact evaluation relying in the use of data. This is because, given that a causal relationship, and researcher use it to make predictions about what would happen if a policy or an action goes one way or the other. The question of cause and effect can be assessed by correctly answering questions of *What is the causal relationship of interest?*, *What is the best way to capture and measure this effect?*, *What is the identification strategy*, and *What is the statistical inference mode that is going to be used?* [4, 3]. Causal effects can be evaluated through the inference of causality, which can be difficult to determine, since the relationship studied may not be so direct and it may be other factors that have the effect. Causal effects can be intuitively defined as the magnitude in which an outcome changes, when a unit of the treatment or policy is changed.

Impact evaluations are needed to make decisions about growth and effectiveness of programs, and adjustment of benefits or requirements to obtain those benefits. This project aims to answer a causal effect question of the effect of conditional transfer program *Avancemos* in the field of secondary school education, by using the innovative approach in this field of Machine Learning modelling. Machine Learning methods can take into account that not all real-world data are unstructured and the effect of interventions can often be observed, for example, by stratifying the data collection across multiple environments, the approximation abilities of modern machine learning methods may prove useful to model nonlinear causal relations among large numbers of variables[5].

Education has had a major role in the development of the Costa Rican society. Since the abolishing of the military forces in 1948, the resources assigned to this prescribed institution, were used to strengthen the education and health sectors, leading to an increased level of development with respect of it's neighboring countries[6]. There seems to be a consensus in society that education is an integral part of the country's social contract[7] and the identity of the modern local citizen[8].

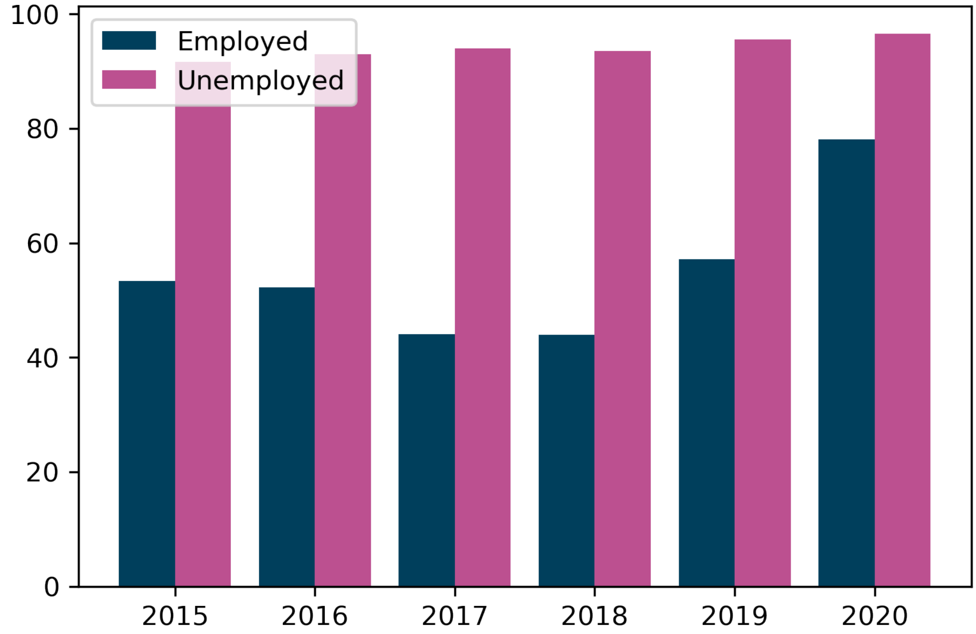
A well educated society is necessary to to constitute a highly qualified workforce, which can lead to significantly improve human development. For the most vulnerable individual, the educational level, specially secondary school education, can be a door for better opportunities such as access to universities and other specialized and advanced technical career developments. A society where only the wealthier members have access to education, is a society headed towards social inequality and the many problems that come with it.

Decisions to stay in school and to work are not necessarily mutually exclusive, however, enough time in school could reduce the time available to work, so child labor is less likely to happen in that situation. Promoting education can be a tool to combat child labor. Costa Rican legislation establishes that that child labor is illegal when the individual is under 15 years of age; after that age it is regulated by the Ministry of Labor and Social Security (*MTSS*), covering self-employment, the formal or informal sector, as well as family work. Child labor should not exceed 6 hours a day and 36 hours a week and this may be limited when the work activity poses a risk to the development of the child in their health, physical, mental and emotional capacities or if it has a negative impact on the minor's attendance at educational center. Most of the cases of child labor are driven, mostly, by the socioeconomic situation of the minors' families, and causes young people to abandon their studies due to the length of working hours, making the decisions to go to school not independent, and competing with each other in opposite directions.[9].

Figure 1.1 shows the percentage of minors from 12 to 17 years old that attend any type of formal education for the years 2015 to 2020. It can be noted that in spite off the percentage of employed attendance has seen a strong raise, specially in 2020, there is still a considerable gap that can be an indicator of the troubling decision that has to be made between getting more education or entering the workforce at a young age.

There are mainly two types of secondary schools in the country, academic and technical schools, and they offer both day and nocturne courses. Figure 1.2 shows the evolution of the level of desertion from secondary school for the years 2014 to 2019 in Costa Rica, by school type, according to the data reported by the Ministry of Public Education of

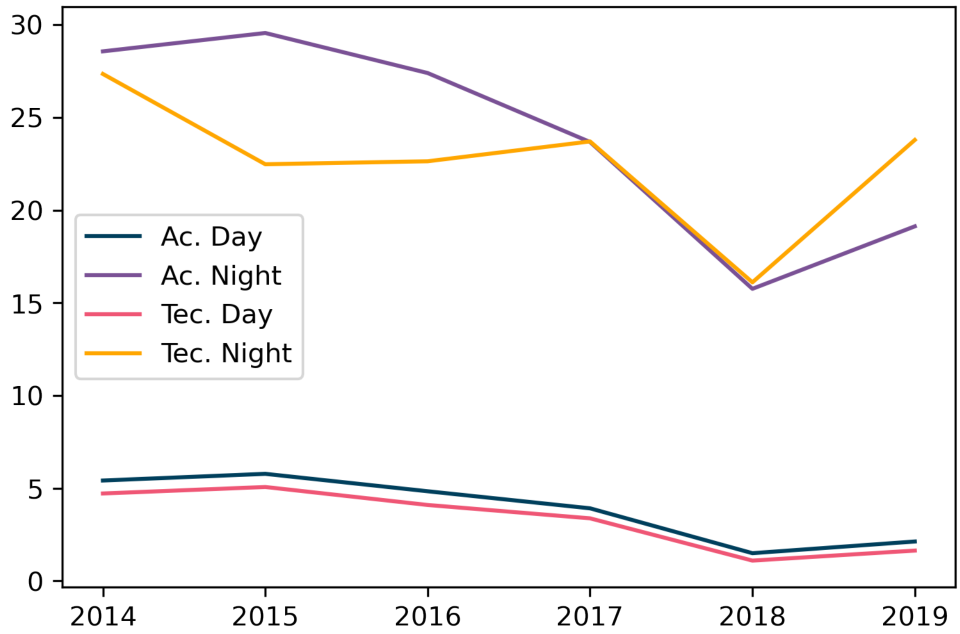
Figure 1.1: Percentage of attendance of minors 12 to 17 years old to formal education by employment condition from 2015 to 2020.



Costa Rica (MEP)[1]. It is to notice that desertion from the system has been in a lowering tendency, but analyzing by between day and nocturnal schools, the latter seem to maintain or even increased, the levels of desertion through the secondary education course.

In the world, and especially in Latin America, since the mid-1990s, it is increasingly common for transfer programs to be used as social policies; whose general objective is to reduce the inter-generational transmission of poverty. $Avancemos$ is a conditional cash transfer (CCT), born in 2006, and has the goal to retain and return secondary-school-age adolescents and youth in Costa Rica's formal education system, [2], or in other words, to prevent the desertion of people from the secondary school system. The $Avancemos$ cash transfer is paid monthly and begins to be delivered once the family has been accepted by the executing entity to receive

Figure 1.2: Percentage of desertion by secondary school type from 2014 to 2019.[1]



the benefit and is maintained during the rest of the year, as long as the student remains in the educational system. Table 1.1 shows the number of beneficiaries of the *Avancemos* Program, which has remained relatively constant throughout the years, with peaks of up to 200K students in the years 2018 and 2019.

Since it's conception *Avancemos* has changed some of the elements that compose the policy, as shown in Figure 1.3. It is to notice that the program has maintained over time the cash transfer component, subject to school attendance, evolving to a simple program with an specific objective, aimed at a population between the ages of 12-25 in the secondary school system, that suffers from poverty and other social exclusion conditions. The latest important change to the policy was in 2013, by adding a second chance for students to repeat classes, before being dismissed by

Table 1.1: Students and families, served by the *Avancemos* Program 2010-2022

Year	Students	Families Represented
2010	185 214	138 415
2011	185 314	139 665
2012	181 570	137 557
2013	171 532	133 147
2014	174 197	136 214
2015	170 774	133 015
2016	168 524	132 260
2017	161 085	126 924
2018	201 503	154 637
2019	200 526	154 353
2020	184 810	145 286

the program, and which was previously only allowed once.

When there are doubts that different sub-populations of the study group, may have experienced a differentiated impact of the policy, then it might be to the presence of heterogeneous treatment effects [3], and it would need special considerations to be addressed.

Machine learning and causal inference methods are currently gaining popularity as as methods that have the potential to improve the estimation of a policy impact [10] or with applications such as predicting individual treatment effects in fields such as healthcare, education and economics [11].

This investigation is divided in 4 chapters. Chapter 1 introduces the basic concepts of policy evaluation, the origin and objectives of the *Avancemos* program in Costa Rica, and the applications of machine learning in policy evaluation.

Chapter 2 starts with the literature review of some evaluations of worldwide known of Conditional Cash Transfer programs such as *Pro-*

Figure 1.3: Changes in *Avancemos*[2]

Category	Creation - Pilot Phase	2007	2008	2009 - 2012	2013 - Present
COMPONENTS					
Transfer	●	●	●	●	●
Saving Incentive	●	●	●	○	○
Support for education and training offering	●	●	●	○	○
CONDITIONALITIES					
Medical evaluation of families	●	●	○	○	○
School attendance	●	●	●	●	●
Optional: participate in volunteer activities	○	○	○	○	●
OTHERS					
Number of class repetitions allowed	1 time	1 time	1 time	1 time	2 times
Age of beneficiaries	13 - 17	12 - 21	12 - 21	12 - 25	12 - 25
Socioeconomic condition	Poverty	Poverty, vulnerability risk and social exclusions	Poverty, vulnerability risk and social exclusions	Poverty, vulnerability risk and social exclusions	Poverty, vulnerability risk and social exclusions

gresa in Mexico, *Familias en Acción* in Colombia, and the *Macedonian CCT Program*, which provide similar experiences to the similar to the *Avancemos* Program in Costa Rica. After this, there is a broad description of the theory behind causal policy evaluation, that includes the Fundamental Problem of Policy Evaluation, the difference between Heterogeneous and Homogeneous Treatment Effects, definitions such as Average Treatment Effects on Treated and Non Treated, and also it is given a great importance to the differentiation between experimental and observational studies, and the ways that each of this resolve the contractual problem. Other concepts studied include, the concept of matching, which is the use of large data sets and statistical techniques to construct the best possible comparison group based on observed characteristics [3].This chapter closes with a revision of diverse studies that have used Machine Learning Techniques to tackle problems similar to what this investigation faces towards an estimation of a causal effect.

Chapter 3, contains the Experiment and Results of the investigation.

This Chapter starts by explaining the dataset used, going through the data pre processing, the imputation of missing values, the treatment identification strategy, the estimation process used, bases on double machine learning, and the presentation of the results.

Chapter 4 contains the conclusions that come from the results, and some considerations of future work of the subject matter.

Chapter 2

Literature Review

In general, evaluations can be defined as the ones that seek to at least three types of questions[3]:

- Descriptive questions: where it is determined what is happening while detailing the context and opinions of the interested party.
- Normative questions: where there is a comparison between the current situation and the desired situation, through the evaluation of different activities and the fulfillment of objectives, and also evaluating the level of inputs, activities or products.
- Questions about cause and effect: In these types of questions, the evaluation focuses on examining results and determining the difference that a certain intervention (or policy) makes on them.

Impact evaluation is as a particular type of evaluation, in which it tries to answer questions about cause and effect. For cases involving people, they are concerned with measuring the impact of changes in well-being that can be attributed to a particular project, program, or policy. The causal effect can be evaluated through the inference of causality, which can be difficult to determine, since the relationship studied may not be so direct and it may be other factors that have the effect. In addition to measuring this causal effect, impact evaluations are necessary to inform policy makers of a range of decisions, from discontinuing inefficient programs to expanding interventions that work, adjusting the benefits of a program, or choosing between several alternative programs[3]. One of the most classical examples of impact evaluation is the work of Rubin (1974) where he defines the basic techniques that would be adopted by a large number impact evaluation professionals[4] to work with either randomized or non randomized studies, including techniques such as randomization and matching, mainly citing the virtues of the randomized trial experiments in the fact that they reduce the problems raised by correlation between predictors of the output and the relevant treatment variable[12], but also making clear that purely randomized studies are difficult to implement in real world, and most of the time researchers will be faced with observational data to work on impact evaluation.

The choice of evaluation method will depend on three broad concerns: the nature of the question to be answered; the type and quality of data available; and the mechanism by which individuals are allocated to the program or receive the policy[13].

2.1 Impact Evaluation of *CCT* policies

In the developing world, and specially in Latin America, it has been a common practice the use of CCT programs since the decade of the 90s with the objective of reducing the inter-generational transfer of poverty. In a broader perspective, these type of programs have been relatively successful, giving a positive impact in development and social inclusion [9]. Each country has adjusted the program to its reality and to the typology of its most vulnerable groups, supported on the basis of breaking the cycle of poverty through the axes of education and health. This as a targeted measure for the poorest groups in each country, but also for the most vulnerable groups within the already vulnerable, such as the poor. In many countries worldwide, Conditional Cash Transfer programs achieve a good performance thanks to adequate monitoring and a good evaluation that shows that the program meets its objectives. The number of actors involved and the need for management comprehensive information to verify compliance with the conditions have interacted in such a way that they promoted a creative development in terms of monitoring and management. For its part, the excellence of the systems and the high degree of transparency of documentation and information that characterizes most programs contributed to the attractiveness of CCTs policies.

2.1.1 Familias en Acción

The program *Familias en Acción* is a Colombian Goverment program targeted to the poorest 20% of the population first implemented in 2001. Within this program, households receive cash if they comply with certain conditions, which include sending children to school and, in the case of young children, taking them to growth and development checkups in the local health centers [14].

Attanasio, Meghir and Vera-Hernández (2004) evaluate the *CCT* program *Familias en Acción* from Colombia, which is a program similar to *Avancemos*, with the objective to increase the accumulation of human capital in rural areas, by improving the nutrition, health and education of families[15]. The treatment studied by the authors is taking part in *Familias en Acción*, which was analyzed when families where given an extra cash transfer to be part of the program and also when they were not given any cash transfer. The target population including families living in one of the Municipalities, where the program is taking place, suffering from poverty conditions with one or more children under the age of 18, and the control group was formed by families with the same characteristics as the treated, but living in municipalities where the program was not offered. Differently from the other CCT programs discussed in this chapter - the evaluation of Familias en Acción doesn't include a random control group. The results of the program show as follows - increase in enrollment rate in the age group 14 - 17 living in urban areas, as well as improvement of nutritional status and compliance with medical check-ups and DTP vaccinations for children between 0 and 6 years old of the rural areas. Familias en acción show a positive impact on female labour supply in the urban areas, which indicate that women are working more,in order to offset the income loss from keeping their children at school.

2.1.2 Progresa

The Programa de Educación, Salud y Alimentación (Progresa) was launched in August 1997 and is part of the targeted social policies or poverty reduction strategies in Mexico. It has the particularity that the designation of the beneficiaries went through a selection process that started from the

detailed knowledge of the localities of the country and the living conditions of the population that resides in them. The objective of Progresa was to support families living in extreme poverty, in addition to enhancing the capacities of its members and expanding their opportunities to achieve better levels of well-being, through actions that would promote the elevation of their living conditions to through the improvement of educational opportunities, health coverage and food. The administration of the program found a total of 506 communities that qualified for the program, and started collecting data from all of these communities, after that, they randomly selected a total of 186 of which conformed the control group, by postponing the implementation of the program for two years. [16]. Progresa program offered cash transfers to randomly selected mothers, subject to taking a role in prenatal care, nutritional monitoring of children, and the children's regular school attendance of more than 85%. The size of the grants increase proportionately to the grade and is moderately larger for girls compared to boys. [17].
The impact of the program resulted high in increasing the enrollment rate among all beneficiaries, as well as school performance, whose improvement "spilled" also among non-eligible boys in the control villages. Progresa proved to successfully diminishing child labour in the poor communities since more children were choosing to stay at school. Because of its positive results and early implementation, the program received significant attention and publicity. Similar models were in fact used to be implemented in CCT programs among other Latin American counties.

Attanasio, Nehir and Santiago (2012) find that is the using observational data and propensity score matching, the program showed to effectively reduce the child labor and to improve educational performance.

2.1.3 Macedonian CCT Program

The Macedonian Conditional Cash Transfer for Secondary School Education is a social protection program of the Macedonian Ministry of Labour and Social Policy, first implemented in the academic year of 2010-2011. The main objective of the program is to increase secondary school enrolment and before all reduce the high drop-out rates among children from the poorest households of the population. In order to benefit from the program - families should also be part of the social financial assistance (SFA) and have children who successfully completed the primary education and are of age for secondary education enrolment. Another requirement is for the students to keep an attendance rate of at least 85%. The total amount of subsidy provided by the *CCT*, if all requirements are met is of USD 240 per year, divided in four equal instalments. The payments are issued in correspondence with the school year quarters - at the end of each quarter the school officers insert the attendance data in the *CCT* system and the Ministry of Labour issues the installments. For the first two years of the program the payments were issued trough printed checks, which can be cashed in banks and post offices, from the third year ahead the installments are directly transferred to the beneficiaries' bank accounts. The impact evaluation focused on understanding two main questions - whether the installments should be paid to the mother or to the household head and if the installments should be equal or there should be a graduation bonus at the end of each school year. To assess half of the municipalities issued the payment to the mother and the other half to the household head (the family member registered in the Social Welfare Center for SFA). Similarly in half of the municipalities instalments were equal and in the other half received smaller initial installments and a larger bonus for completing the school year. All the data was processed together with data about school attendance rates. Examining the short- and medium-term impacts of the Macedonian *CCT* program there is a noticeable impact on school enrollment rates for children of secondary school age. There isn't a substantial difference when it comes to school attendance, which was already high. When it comes to household or child outcomes of receiving the cash transfer payments in equal instalments or receiving a higher final installment as a graduation bonus - there aren't any significant differences. There is however strong impacts of paying the mother versus the head of the household when it comes to food expenditure at the end of the second year of the Macedonian *CCT* program. This may be an indication of change in the level of women empowerment in their households. Starting from the fourth year, after acknowledging the results of the first evaluation of the *CCT* program, the randomization of payment modalities was removed by the Ministry of Labour and Social Policy and installments were only directed to the mothers[18].

2.2 The fundamental problem of policy evaluation

Assessing the impact of a policy is not an easy task, what is sought is to determine what would have happened if had not been implemented, so that it can properly measure the impact of the same, which can be seen as the difference between the situation with and without politics. Since it is not possible to know exactly what would have happened in the scenario without politics, there are different methodological strategies that allow estimating what would have happened to the variables in that situation. Since it is not possible to observe the same individual in both the situation where the individual received the treatment and the situation in which the individual did not, then it is needed to build a counterfactual[3], which is defined as the situation of what would have happened to the treated individual in the absence of treatment, can also be seen as what would have happened to the untreated individuals if they had been given the treatment. Policy evaluation methods are interest in establishing causal relationships of a policy with certain desired objectives, to better illustrate this, suppose it is desired to measure the impact of the policy T on the outcome variable Y.

First assumption, that is going to be made is that the T will denote only two values, 1 or 0. If $T = 1$, it means that the treatment was received, if $T = 0$ means that the treatment was not received. Then, it will be assumed that there is N amount of individuals in a population and that i represents a single individual in this population. Therefore, the outcome of receiving the treatment can be denoted as follows as $Y_i(T = 1)$, which can be alternatively expressed as $Y_i(1)$, and similarly, the outcome of not receiving the treatment, $T = 0$, can be expressed as $Y_i(0)$.

So, the potential outcome and individual can be expressed as follows[3, 13]:

$$Y_i(T_i) = Y_i(1)T_i + Y_i(0)[1 - T_i] \tag{2.1}$$

Which can be rewritten as:

$$Y_i(T_i) = Y_i(0) + T_i[Y_i(1) - Y_i(0)] \tag{2.2}$$

Where the first part $Y_i(0)$, denotes the outcome of not being treated, and the second part $[Y_i(1) - Y_i(0)]$, which will be called, which will be later recalled in Equation 2.3, denotes the difference of being treated vs not, or in other words, it is the difference in outcome between participating and not participating on the policy. It is of interested to measure these two outcomes for the same individual, at the same precise moment in time, but given that these are two mutually exclusive states, which are called *counterfactual*. This measuring task becomes impossible to do, since if we assume that individual i took that treatmentment ($Ti = 1$, it is already known that it is impossible to find the same individual i not taking the treatment at the same moment of time. This problem is known as the as the *Counterfactual Problem or Fundamental Problems of Policy Evaluation* [12, 3, 17, 13].

$$\Delta Y_i = Y_i(1) - Y_i(0) \tag{2.3}$$

So, recalling Equation 2.3, and taking into account that there is no solution to this problem, what one can do would be to estimate a solution to this by finding not solving the problem individually, but finding a sample of treated and untreated individuals and comparing their outcomes in average.

One key solution for the estimation and resolution of this counterfactual program is that instead of trying to estimate on an individual level, moving to a group level, even though there will no be a perfect pair for every single individual that took the treatment, some statistical properties will help to generate two groups of units that, if their numbers are large enough, are statistically indistinguishable from each other at the group level[3]. This groups will be called *treatment group* and *control group*. Finding a good control group to compare with the treatment is a crucial task to do, because if the comparison is not got the it will not accurately estimate the true counterfactual, then the estimated impact of the program will statistical biased [3, 4].

Another key aspect to take into consideration when evaluating a policy on a group is if the policy affects everybody the same, so we can imply the policy has an homogeneous effect or if a policy affects individuals differently, where we can say the policy has an heterogeneous effect [3]. Now, lets suppose that α represents the average outcome without treatment. We can rewrite then the potential outcomes for the treated individuals and non treated individuals as in 2.4 and 2.5, respectively:

$$Y_i(1) = \alpha + \beta_i + \epsilon_i \tag{2.4}$$

$$Y_i(0) = \alpha + \epsilon_i \tag{2.5}$$

And where ϵ_i refers to a non observable component of the outcome Y. And if we replace this equations in 2.3, with the definition of heterogeneous and homogeneous effects we get the following:

$$Y_i(1) - Y_i(0) = \beta \ \forall i \tag{2.6}$$

$$Y_i(1) - Y_i(0) = \beta_i \tag{2.7}$$

Where Equation 2.6 represents to the homogeneous effects, since β is the same for every i, and Equation 2.7 the heterogeneous effects. Getting back to the estimation of the effect, it was established that the resolution of the *counterfactual problem* would be resolved by the comparison between the outcomes of subsamples of a population that received the same treatment, and a control group, this is crucial since Constructing a counterfactual in a convincing way is a key ingredient of any serious evaluation method[13]. If we made a naive comparison of averages between treated and non treated to

Treatment effects are usually estimated [3, 4] by the calculation of the *Average Treatment Effect (ATE)* (see Equation 2.8), the *Average Treatment on the Treated (ATT)* (see Equation 2.9, the *Average Treatment on the Non Treated (ATNT* (see Equation 2.10, and the *Local Average Treatment Effect(LATE)* (see Equation 2.11).

The ATE reflects the average effect of the treatment in the whole population. It does not matter if the individual received it or not, and is also the expected effect of giving the treatment to a random chosen individual:

$$ATE = E[Y_i(1) - Y_i(0)] = E[\beta_i] \tag{2.8}$$

ATT shows the average effect of the individuals that took the treatment, either by selection or by participating in the treatment:

$$ATT = E[Y_i(1) - Y_i(0)|T_i = 1] = E[\beta_i|T_i = 1] \tag{2.9}$$

$ATNT$ shows the average effects on individuals who did not were part of the treatment, either by being selected or not wanting to participate:

$$ATNT = E[Y_i(1) - Y_i(0)|T_i = 0] = E[\beta_i|T_i = 0] \tag{2.10}$$

$LATE$ is similar to ATE in that is the average effect, independently of having been given the treatment or not, but in a sub group of the population that share certain common characteristic.

$$LATE = E[Y_i(1) - Y_i(0) | i \in \{\cdots\}] = E[\beta_i | T_i = 1] \tag{2.11}$$

Which under the assumption of homogeneous treatment effects ($\beta_i = \beta \ \ \forall i$) will lead to Equation 2.12:

$$ATE = E[\beta_i] = ATT = E[\beta_i | T_i = 1] = ATNT = E[\beta_i | T_i = 0] = \beta \tag{2.12}$$

Still the problem continues to be the counterfactual problem, because it is not possible to observe it in this averages as it can be summarized in Table 2.1:

Table 2.1: The missing conterfactual problem

Treatment Effect	Specification	Non Observable Conterfactual
ATE	$E[Y_i(1) - Y_i(0)]$ $= ATT\ P(T=1) + ATNT\ P(T=0)$ $= E[Y_i(1)] - E[Y_i(0) \mid T_i = 1]\ P(T=1) + E[Y_i(1)] - E[Y_i(0) \mid T_i = 1]\ P(T=0)$	$E[Y_i(0) \mid T_i = 1]$, and $E[Y_i(0) \mid T_i = 1]$
ATT	$E[Y_i(1)] - E[Y_i(0) \mid T_i = 1]$	$E[Y_i(0) \mid T_i = 1]$
$ATNT$	$E[Y_i(1)] - E[Y_i(0) \mid T_i = 1]$	$E[Y_i(0) \mid T_i = 1]$

In spite of not being capable of observing an individual overcome treatment and not overcoming treatment at the same time, perhaps it is possible to compare a sample of individuals in a group that received treatment and in a group that was untreated and compare their outcomes in a *naïve* comparisson. The potential outcomes for the treatment group will be:

$$E[Y_i | T_i = 1] = E[T_i Y_i(1) + (1 - T_i) Y_i(0) | T_i = 1] = E[Y_i(1) | T_i = 1] \tag{2.13}$$

And for the non treated group will be:

$$E[Y_i|T_i = 0] = E[T_iY_i(1)+(1-T_i)Y_i(0)|T_i = 0] = E[Y_i(0)|T_i = 0] \quad (2.14)$$

Note that also from Table 2.1, Equation 2.13 and Equation 2.14, are not in the non observable column, so they can be estimated directly from the data. The difference between the expected values will be then the following:

$$E[Y_i|T_i = 1] - E[Y_i|T_i = 0] = E[Y_i|\widehat{T_i = 1}] - E[Y_i|\widehat{T_i = 0}] \quad (2.15)$$

But, what if an individual does not want to participate in the treatment, or on the contrary, is more motivated than others to participate on ir? Selection bias occurs when the reasons for which an individual participates in a program are correlated with outcomes. Ensuring that the estimated impact controls for selection bias is one of the major objectives and challenges for any causality estimation[3]. Going back to our potential outcome in Equation 2.1, and adding in Equation 2.4 and 2.5, we get the following expression:

$$Y_i(T_i) = \alpha + \beta_i T_i + \epsilon_i \quad (2.16)$$

Now, from 2.8, assume that $\beta = \beta^{ATE}$, and that we substract and add this term multiplied by the treatment term T into Equation 2.16, getting the following expression:

$$Y_i(T_i) = \alpha + \beta_i T_i + \epsilon_i = 1 \quad (2.17)$$

2.2.1 Randomized experiments vs Observational Data

Random experiments, mostly, are time consuming, expensive, and may not always be practical.[17], that is why, investigation often turns into observational studies, which as the name implies, lack the experimental random assignation of the policy, and present a different series of challenges of fundamental difficulty in estimating treatment effects. Researchers tend to seek help in the idea of propensity score methodology, which summarizes a large number of control variables in a single variable, known as the *propensity score*, and then to use this to establish a control group to contrast with the group that received the treatment when random assignation was not possible. This is called a *pseudo experiment*. Matching techniques have been used to evaluate development programs in a wide array of settings. They can be often combined matching with other methods to produce more robust results. They consist on finding a non treatment unit for each of the treated that has the most similar characteristics possible, and are best used when the program that is evaluated was not randomly assigned. The set of non treated individuals more alike to the treated ones can be used as a control group to estimate the counterfactual [3, 4].

2.2.2 Dealing with heterogeneous effects

The distinction between homogeneous and heterogeneous responses is central to understand what parameters alternative evaluation methods measure. In the homogeneous linear model, common in elementary econometrics, there is only one impact of the program and it is one that would be common to all participants and nonparticipants alike. In the heterogeneous model, the treated and non-treated may benefit differently from program participation. In this case, the average treatment effects among the treated will differ from the average value overall or on the untreated individuals. Indeed, we can define a whole distribution of the treatment effects[13]. Mingliang and Tobias study the estimation of the causal impact of an endogenous treatment variable in the context of a non randomly assigned treatment, but selected by the individual[19], treating this issue with the use of the *Bayesian Approach* by taking we consider Bayesian estimation in a variant of the correlated random coefficient *CRC* model, and by establishing a tree equation system where and observed scalar represents the endogenous treatment variable and an individual causal effect parameter. Efforts in this regard primarily have either focused on the Markov chain Monte Carlo *(MCMC)* implementation or have focused on problems associated with weak instruments generally, has discussed priors that yield posteriors similar to sampling distributions for the two-stage least squares, and limited information maximum likelihood estimators[19]. The model considered by the authors consists on the following equations:

$$y_i = \beta_0 + x_i\beta + s_i\theta + u_i \tag{2.18}$$

$$s_i = \delta_0 + x_i\delta + z_i\gamma + \theta_i\rho + s_i\theta + v_i \tag{2.19}$$

$$\theta_i = \eta_0 + x_i\eta + w_i\lambda + \epsilon_i \tag{2.20}$$

Where, Equation 2.18 is the equation of the outcome of interest, and s_i is the endogenous treatment variable. z_i and w_i are two instrumental variables, x_i is common in all the equations, and represents the covariates. The parameter θ_i denotes that the treatment can have effects that vary individuals. Equation 2.20 is important because it denotes that an unobservable return θ can depend on other non related characteristics. It can also be seen as a non traditional instrument versus z_i which is an instrument in a more traditional way. Wooldridge on the other hand provides a more general approach to the problem[20]

2.3 Impact evaluation using machine learning methods

2.3.1 Supervised Learning vs Causal Learning

Most Machine Learning focus on predicting outcomes rather than understanding causality. *Supervised learning* is the most common training paradigm for performing classification tasks and prediction applications. In this approach, the learning algorithm adaptively adjusts the difference between the desired output and the one generated by the neural network. Thus, the network learns to automatically perform a prescribed task. This requires the network designer to know in advance the response of the network in a sample of cases that are used for training. In this case, as previously stated, all inputs are specified, as well as outputs. For each example proposed in the network, the outputs they produce are compared with the desired outputs. Before the entire subset of training samples had been processed, the network connection weights had to be updated. The training or updating of weights is done in such a way that the measure of the error is reduced.

Table 2.2: Causal Learning vs Supervised Learning

	Supervised Learning	**Casual Inference**
Predicts	Outcome $P(Y\|X)$	Effect of change $P(Y\|doX)$
Assumption	Passive observer	Decision maker
Train test	Equally distributed	Distribution shift
Validation	Easy, via hold-out	Fundamental challenge. Better prediction is NOT better casual estimation
Feature set	Quantitative (over fit / under fit)	Qualitative - could cause a bias in the estimate
Domain Knowledge	Nice to have, deep neural network are doing beyond humans without	Essential to make assumptions to avoid pitfalls
For Who?		

The selection bias is a problem that occurs when assigning individuals to a control group, it brings a distortion in the statistical analysis and the inference that is made from this. If the selection bias is not controlled and is too high, estimation of the effects of a policy can be shown partially the effects or not shown at all[4].

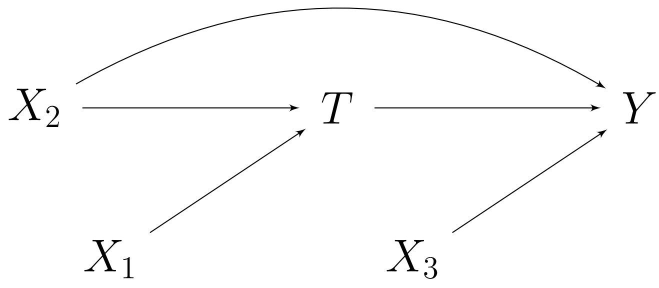
2.3.2 Selection of covariates in high dimensional spaces

Generally, researches face a problem with selection of confounding variables, since we want to identify the optimal model, select those predictor variables that are most appropriate to include in our model. A large inclusion of variables can lead to the curse of dimensionality, where the data becomes sparse because of the large number of observations and predictions can turn out wrong. The solution to this problem could be to choose the appropriate features needed for the model, using techniques such as *Principal Component Analysis*, The problem is when the number of candidate predictor variables is large, because then the number of possible regression models will also be large. Asharian, et al. [21], note that there are several types of predictors, which are represented in Figure 2.1, which are:

- X_1 which are predictors that are only related to the treatment.
- X_2 which are predictors that are related both to the outcome and to the treatment.
- X_3 that are only related to the outcome Y

A practical problem empirical researchers face when trying to estimate treatment effects is deciding what conditioning variables to include. When the treatment variable or instrument is not randomly assigned, a researcher must choose what needs to be conditioned on to make the argument that the instrument or treatment is exogenous plausible.[22]. Belloni et. al. also note that given the cost of producing random experiments and random treatment assignment, researches have to make adopt

Figure 2.1: Types of predictors



quasi experimental approaches, to some problems. One approach could to identify if there is some other external variable, such as eligibility for receiving certain benefit from a government program or service, that is either randomly assigned or the researcher is willing to take as exogenous conditional on the right set of control variables, or to rely on matching techniques which also introduce a another problem that is the choosing of the variables that will be used for the matching [22].

Current estimation methods that exploit approximate sparsity employ different types of regularization aimed at producing estimators that theoretically perform well in highdimensional settings while remaining computationally tractable. Many widely used methods are based on $L_1 - penalization$. The Lasso method is one such commonly used approach that adds a penalty for the weighted sum of the absolute values of the model parameters to the usual objective function of an M-estimatorMany authors agree that using the Lasso Regression and variations can be beneficial on the t Belloni, et al. (2011) use a variation of the Lasso regression to correct this called [23]. Lasso (l1-)penalties are useful for fitting a wide variety of models. Newly developed computational algorithms allow application of these models to large data sets, exploiting sparsity for both statistical and computation gains [24]

2.3.3 Dealing with missing values in observational data

Dealing with missing values is a common thing when talking about observational data. These values are those that do not appear due to any event, such as errors in the transcription of the data or the lack of willingness to answer certain questions in a survey. Data may be missing randomly or non-randomly. Simply discarding random missing data can disrupt data analysis as they decrease sample sizes and thus the power of hypothesis testing. Non-random missing data also cause a decrease in the representativeness of the sample, and can be a problem even more severe for populations that are already underrepresented for example in surveys. [25, 26, 27].

Following [28] in order to insert sensible values for missing data we must rely on some model relating unobserved values to observed values. Hence, I see the best practical approach to be one where we can insert more than one value for a missing datum in a way that reflects our uncertainty; the inserted values should reflect variation within a model as well as variation due to a variety of reasonable models.

2.4 Double Machine Learning and Sample Splitting

Machine learning methods seem to be effective in prediction contexts, however, good performance in prediction does not necessarily translate into a good estimation in inference about causal parameters. Double Machine Learning (*DML*), also known as *Orthogonalized Machine learning* combined with sample splitting can help construct high quality point interval estimates of the causal parameters. Sample splitting and the application of cross-fitting is a central part of Double Machine Learning (*DML*). These estimators rely on two robust estimators [29] [] for example a propensity score and a outcome estimator. The bias of the treatment effect estimator is like a product of the bias of the two estimation, that means that if one of the estimators is unbiassed, then the bias of the treatment effect will also be unbiased.

Problems with this approach are the we are introducing some kind of overfitting, because we are using the own observation bias. sample splitting which is training the machine with one half of the data and evaluating the effect estimator with the other half can be used to reduce this bias. (cite Zenh and van der Laan 2011 ,belloni et al 2011 [29] but can produce residual cofounding bias, this method also allows one predictor to converge fast and other to converge slowly.

Machine Learning can be carefully integrated into causal framework for statistical inference, for example in modern data ecosystems where we often have a great amount of confounding variables, taking into account that we rarely have the knowledge to specify an *a priori* correct parametric regression, ML algorithms must be carefully integrated within a formal framework for causal and statistical inference.

Chapter 3

Experiment and Results

3.1 Database description

The National Household Survey *(ENAHO)*, is an annual survey since the year 2010 in Costa Rica. Consists of a data collection program whose focus is associated with the level of well being of the population, especially focused on the composition of household income, its distribution and household characteristics and the population living in them. It also includes the study of housing ownership and its characteristics, people's access to education and social security, as well as the working population and the conditions of these jobs, among others. Likewise, each year specific investigations or special modules are carried out, within which access to social programs, use of health services, telecommunications, child labor, migration, citizen security, and breastfeeding can be mentioned. The Survey is carried out in July of each year, and allows information to be obtained at the national, regional and urban and rural levels[30]. The database contains all the information corresponding to the dwellings from a representative sample at national and regional level, as well as information on all the households and people residing in these dwellings. It consists in cross-sectional data collected annually, but the same individual is not followed over time, so it is possible to resort to methodological strategies that allow a panel to be approximated, so that it is controlled by groups of individuals who maintain similar characteristics. Each dwelling has an expansion factor, which is defined as weight that is applied to each study unit in the sample to obtain a population estimate, and is interpreted as the number of units in the population that each unit in the sample represents, whether dwellings, households, or persons. For the purpose of this study, weights will not be applied. Table 3.1 shows the amount of households, and persons that are part of this households for the years 2010 to 2020. A downward trend in the average size of the household is noted, as well a huge decline in the numbers for 2020. The downward trend can be explained by the wear of the Survey's sampling frame, which was built alongside the 2011 National Census, a decline in the fertility rates over time, and for 2020, the effect of the COVID-19 pandemic[30]. 12-17 year old's, which are the targets for *Avancemos*, behave similarly, and also the representation of the participants of the program, but this is only in the representation of the survey without taking the weighting, because as it was seen from Table 1.1, the actual participants do not decrease as much.

Table 3.1: Number of households and persons in the National House Survey and recipients of *Avancemos*

Year	Households	Persons	12-17 years old	In Avancemos
2010	11 603	41 184	5 071	1 062
2011	11 721	40 860	4 979	1 049
2012	11 373	39 390	4 597	1 172
2013	11 219	38 779	4 469	1 189
2014	11 495	38 299	4 189	1 094
2015	11 277	37 291	3 706	843
2016	11 335	37 006	3 686	853
2017	10 712	34 843	3 508	848
2018	10 942	35 096	3 387	898
2019	11 006	34 863	3 255	796
2020	8 124	25 530	2 450	661
Total	**120 718**	**403 241**	**43 297**	**10 465**

3.2 Data Processing

Even though the Survey is passed every year, there are significant differences in the number of questions asked across time, where the most common variation through years is the addition of new questions, and even entire new sections, but also it is possible to find discrepancies in questions, such as questions that were deleted in following years, questions included new or different categories in different years, and there were a few cases of questions that even though had the same column name, corresponded to a different question entirely. Table 3.2, shows the total columns present in the Survey by year. Most of the discrepancies can be found following the years 2011, 2013 and 2016, meanwhile in the rest of the years, it is usual to find just new questions added. Table 3.2 also shows a total of 377 common columns, which was reached constructed by intersection of the common columns throughout the years meanwhile controlling the aforementioned discrepancies, either by re-codification of columns, renaming, or not taking them into account.

Table 3.2: Number of Questions in *ENAHO* and common columns

	Year										
	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**
Number of Columns	485	510	523	521	539	563	576	579	571	572	625
Common Columns						**377**					
Common Columns with less than 50% missing values						**131**					

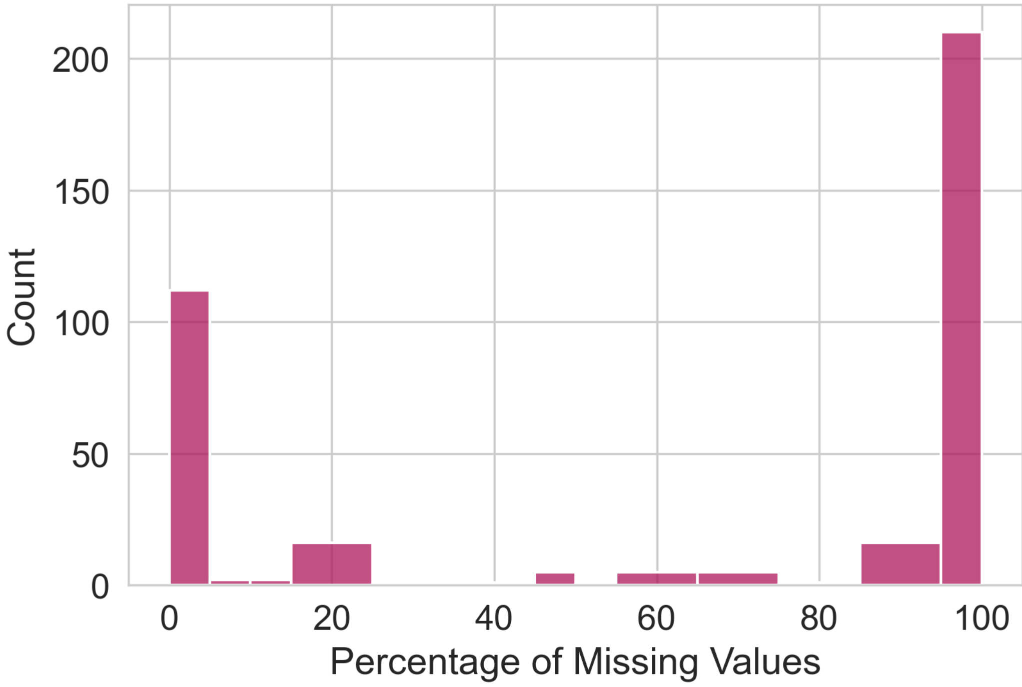
3.2.1 Imputation of Missing Values

One of the most common problems that we can find when working with a data set is the existence of records with null values. Since we are dealing with a population that may be living marginalised, and in poverty, it may be necessary to impute a value to missing value rows in order to use them in a subsequent analysis. Figure 3.1 shows a distribution of frequencies of the percentage of missing values from the common columns the data found in Table 3.2, where it is possible to see there is an important number of columns which are missing all their values, which probably account for questions related to the head of the household.

Figure 3.2, shows the same distribution but after discarding columns with more than half of their rows missing. The imputation of values was done. From the total of 131 columns (see Table 3.2), 114 correspond to categorical variables, and 17 to numeric variables. The missing values were imputed differently according to their types as follows:

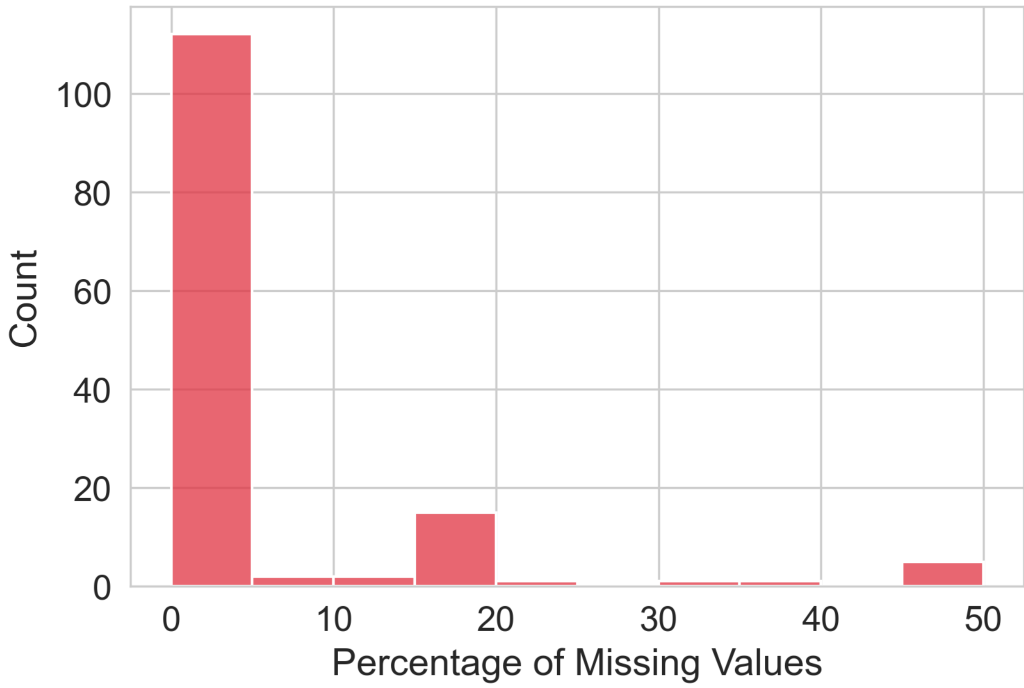
For each column, a estimation using the Bayesian Ridge method was performed, using the column each column with missing values once as the output variable, and the rest of the columns as predictors. In general, when fitting a curve using Bayesian ridge regression, the selection of the

Figure 3.1: Missing Values from Common Columns



initial values of the regularization parameters can be important, this is because the regularization parameters are determined by an iterative procedure that depends on the initial values, that is why this was done iteratively, starting from the columns with less missing values, to the columns with more. The chosen estimator for this was a Bayesian Ridge Regression since this type of linear regression uses probability distributors instead of point estimates, which makes it specially effective with data that is far from normally distributed, by making the assumption that the data target data is distributed around the Guassian distribution. To simplify calculations, only the square root of the total columns which is around 12, were used to calculate each missing value. After this, for the numeric columns, the median of the nearest values was imputed as missing value, and for categorical, the most frequent value of the nearest values was imputed.

Figure 3.2: Missing Values from Columns with



3.3 Identification Strategy

Initially, by only looking at a mean comparison as shown in Table3.3, we can observe that there is statistical significant difference between the amount of young people working a according to being recipients or not to the CCT from *Avancemos*. The percentage of workers between the non recipients is larger than from recipients.

Table 3.3: Descriptive Statistics: Means of the outcome variable for the total sample, treated and non treated groups

Variable	Total of the Sample	With *Avancemos*	Without *Avancemos*	Difference	$P > \|t\|$
Percent. Working	5.68	2.89	6.62	3.79	0.0000***
Observations	43 297	10 936	32 361		

But with this simple comparison, the only think it can be accounted for is the correlation between variables, and as it was explained in Chapter 2, this can not be taken as a causality, because there could be external factors that could be affecting these individuals, such as household income, the physical conditions of the dwelling, and other that we are not taking into account at the moment. It is important also to delimit the access or qualification to *Avancemos*, since, besides the age restrictions, it is targeted at persons living under poverty, and this is something that can be known from the database. Figure 3.4, shows the distribution of the 12-17 year olds, by their working status, their poverty level and if they are part of *Avancemos*.

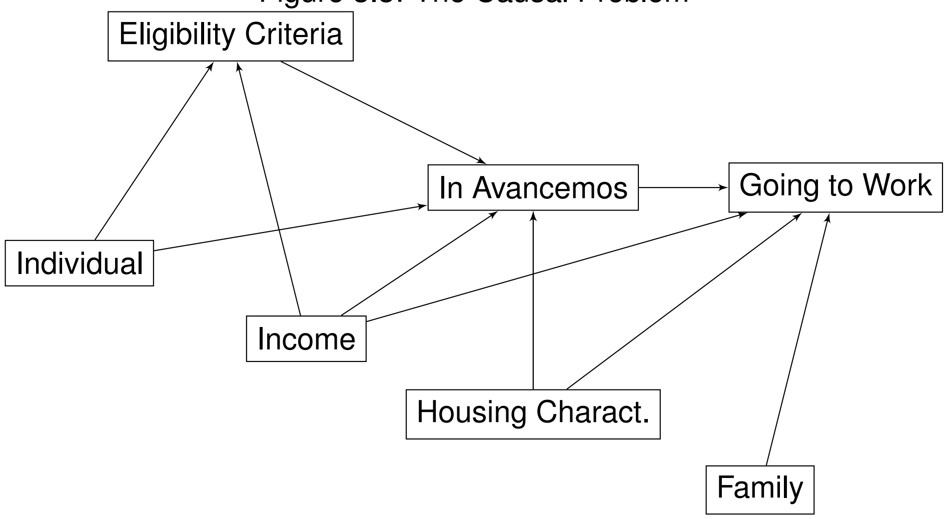
Figure 3.3 illustrates how the relationship in a causal graph between the different groups of variables. This introduces also the idea behind that some variables may affect both the treatment and outcome at the same time, and that is why it is important try to conserve this interactions. As it has been stated, our treatment variable for this investigation is a dummy variable that corresponds to a value 1, if and individual par-

Table 3.4: Distribution of people eligible for *Avancemos* by age and by poverty status

in ***Avancemos*?**	**In Poverty?**	**Working?**	**Total**
No	Yes	No	9 284
No	No	No	20 923
Yes	Yes	No	4 867
Yes	No	Yes	197
Yes	Yes	Yes	119
No	Yes	Yes	738

ticipated in *Avancemos*, and 0 if not. The outcome variable is a dummy variable that states 1, if the person works and 0 if not.

Figure 3.3: The Causal Problem



From Figure 3.3 in a partially linear regression setting in *DML*, we can define the investigation following equations of interest as the Outcome Equation and the Treatment Equation as follows:

$$Outcome : Y_i = \beta_0 t_i + g_0(Z_i) + u_i \; E(u_i|Z_i, t_i) = 0 \tag{3.1}$$

$$Outcome : t_i = m_0(Z_i) + v_i, \quad E(v_i|Z_i) = 0 \tag{3.2}$$

Where,

- Y_i is the outcome variable, denotes 1 if the individual i works, and 0 if not.

- t_i is the treatment variable, denotes 1 if the individual 1 recieved *Avancemos*, and 0 if not.

- $Z_i = (z_{1,i}, \cdots, z_{p,i})$ is a vector of controls or cofounders, with the variables related to the individual characteristics, household, income and family.

- $m(.)$ and $g(.)$ are nuisance functions.

- β_0 is our parameter of interest, and will denote the effect of the treatment.

- We assume that t_i is exogenous, conditional on $g_0(Z_i)$, this asumption is holded by controlling by the Eligibility Criteria.

- The function $g_0(Z_i)$ captures the relation between the nuisance parameters and the outcome.

- The function $m_0(Z_i)$ captures the relationship between the nuisance parameters and the treatment.

The method that will be utilized then does the following:

- Use a Machine Learning method to predict the outcome. In this case several $sklearn$[31] predictors will be used ($Lasso$, $Decision\ Trees$,$Random\ Forest$):

$$Y \text{ from } z \rightarrow u = y \hat{-} \hat{g}_0(Z)$$

- Use a Machine Learning method to predict the treatment. Same as the last step, a variety of methods will be utilized.

$$t \text{ from } z \to v = t - \hat{\hat{g}}_0(Z)$$

- Regress residuals on residuals to get the treatment effect:

$$y - \hat{y} \text{ on } t - \hat{t}$$

If we consider a treatment $t \in \{0, 1\}$, and Equation 3.1 and Equation 3.2, the *Average Treatment Effect (ATE)* is:

$$\hat{\beta}_0 = E\left[g_0(1, Z) - g_0(0, Z)\right] \quad (3.3)$$

And the *Average Effect of the Treatment on the Treated (ATT)* is:

$$\hat{\beta}_0 = E\left[g_0(1, Z) - g_0(0, Z) | t = 1\right] \quad (3.4)$$

The controls or cofounders, affect the treatment coefficient via the nuisance functions. These functions can be complicated to estimate [29] but this *DML* can be used to learn them.

3.3.1 Hyper Parameter Tunning

Parameter Hyper-tunning was done to improve the estimation of the different learners used. Paramater hyper tunning is a tool used to automate the process of combines different parameters of an algorithm, The processes are really nothing more than choosing a series of values for each hyperparameter, making the possible combinations and starting to try different values. Some methods perform an exhaustive search, and others choose only certain items from the search space. They are very effective methods and that in general allow us to obtain the optimal hyperparameters. Also there is also a chance for tuning the *DML* by the specification of different levels of $n - folds$ or $n - reps$ which account specifically for the amount folds done in sample splitting and the later accounts for the repetition of the sampling. So for example, n_fold= 5, n_reps= 2, means that the model will split the sample in five parts, calculate a model for each part, and repeat that twice. Also, it will alternate the sample that used to calculate $m(.)$ with $g(.)$ and average them.

3.4 Estimation Results

For estimation purposes, four different variation of models were used. First, it was the Lasso estimation, then a Tree Model and two other Ensemble Methods are meta-algorithms that that try to improve the capabilities of learning algorithms in order for them to generate better predictors without having to change the current set of training data.

Table 3.5: Estimated effect of *Avancemos* on child labor

	Lasso	Reg. Tree	Random Forest	AdaBoost
Partially Linear Model n_folds = 5 , n_reps = 1				
	-0.043476	-0.012028	-0.014616	-0.000438
	[0.002402]	[0.001171]	[0.001473]	[0.000519]
	(0.0000)	(0.0000)	(0.0000)	(0.398139)
Partially Linear Model n_folds = 3 , n_reps = 1				
	-0.043464	-0.012479	-0.01579	-0.000015
	[0.002401]	0.001222	[0.001543]	[0.000255]
	(0.0000)	(0.0000)	(0.0000)	(0.954307)
Partially Linear Model n_folds = 3 , n_reps = 2				
	-0.043434	-0.012232	-0.014583	-
	[0.002402]	[0.001218]	[0.001499]	-
	(0.0000)	(0.0000)	(0.0000)	-
Partially Linear Model n_folds = 5 , n_reps = 2				
	-0.043462	-0.012219	-0.013938	-
	[0.002402]	(0.001216)	[0.001486]	-
	(0.0000)	(0.0000)	(0.0000)	-

Adding randomization and then combining predictors has an average effect on the performance of the learning algorithm, which corresponds to reducing variance and increasing bias. From Table 3.5, for each of the four models for the results of the estimation can be seen. For the four models estimated, the *Adaboost* model was the least performing one, perhaps for a bad specification or the large amount of parameters in the model. For the other models, they seem to be coincident on the positive effect of the *Avancemos* Program, but not so much in the magnitude of the coefficients, which are larger in the *Lasso* model vrs, the *RegressionTree* and the *RandomForest Regressor*. The *Lasso* model is working better because the large values of the penalty parameter λ, lets the

Chapter 4

Conclusions and Final Remarks

The *Avancemos* conditional cash transfer program implemented since 2006, whose main objective is to reduce school dropouts and encourage reintegration into the educational system, had a moderate effect on the reduction of child labor during the period 2010 and 2020 .

To evaluate the impact of a social policy such as Avancemos in this case, a novel methodological strategy was used, which allows estimating the effect of the program and taking into account a large amount of information contained in the National Household Survey (*ENAHO*) database.

It is not well known how to choose how to choose the estimator for the nuisance parameters in the Double Machine Learning Method. The large amount of variables used in the model perhaps signals that the Lasso and l_1 *penalized* models could probably deal better with sparse matrix estimation problems, but this is not that clear in the causal machine learning field.

The Social Programs of Conditional Monetary Transfers have become, in Costa Rica and in many countries, specially in Latin America, part of the social policy par excellence. The effective evaluation, which adapts to the new times, will allow a better use of resources and effectiveness in the administration of these.

From the estimation of the effect, even though the effect is low, the percentage of children 12-17 years old who are most vulnerable living in poverty and have to work is relatively low, so a consolidated program such as *Avancemos* is also expected to retain children from leaving the school system, at least until they are of age.

Bibliography

[1] M. d. E. P. d. C. R. MEP, "Indicadores del sistema educativo costarrricense." `https://www.mep.go.cr/indicadores_edu/indice_indicadores.html`, 2022. Accessed: 2022-07-05.

[2] K. Hernández-Romero, "How does *Avancemos* work? best practices in the implementation of conditional cash transfer programs in Latin America and the Caribbean," tech. rep., Inter-American Development Bank, 2016.

[3] P. J. Gertler, S. Martinez, P. Premand, and L. B. R. C. M. Vermeersch, *Impact Evaluation in Practice.* Washington DC, USA: World Bank, 2 ed., 2016.

[4] J. D. Angrist and J.-S. Pischke, *Mostly Harmless Econometrics.* New Yersey, USA: Princeton University Press, 2009.

[5] B. Schölkopf, F. Locatello, S. Bauer, N. R. Ke, N. Kalchbrenner, A. Goyal, and Y. Bengio, "Toward causal representation learning," *Proceedings of the IEEE*, vol. 109, no. 5, pp. 612–634, 2021.

[6] M. Muñoz, "Costa Rica: La abolición del ejército y la construcción de la paz regional," *Historia y comunicación social*, 2014.

[7] OECD, *OECD Economic Surveys: Costa Rica 2018.* Paris, France: OECD- Economic Developtment Review Comittee, 2018.

[8] D. Díaz-Arias, "Reforma sin alianza, discursos transformados, interés electoral, triunfos dudosos," in *Serie de Cuadernos de Historia de las Instituciones de Costa Rica* (University of Costa Rica, ed.), San José, Costa Rica: University of Costa Rica, 2007.

[9] G. Lang, J. Robalino, and B. Soto, "Efecto del programa avancemos sobre el trabajo infantil en costa rica 2005-2009," Master's thesis, Universidad de Costa Rica, Costa Rica, 7 2015.

[10] P. R. Hahn, J. S. Murray, and C. M. Carvalho, "Bayesian regression tree models for causal inference: Regularization, confounding, and heterogeneous effects," *Bayesian analysis*, 2020.

[11] U. Shalit, F. D. Johansson, and D. Sontag, "Estimating individual treatment effect: generalization bounds and algorithms," 2017.

[12] D. B. Rubin, "Estimating causal effects of treament in randomized and nonrandomized studies," *Journal of Educational Psychology*, vol. 66, no. 5, pp. 688–701, 1974.

[13] R. Blundell and M. C. Dias, "Alternative approaches to evaluation in empirical microeconomics," *The Journal of Human Resources*, vol. 44, no. 3, 2009.

[14] O. P. Attanasio, "The Determinants of Human Capital Formation During the Early Years of Life: Theory, Measurement, and Policies," *Journal of the European Economic Association*, vol. 13, pp. 949–997, 12 2015.

[15] O. P. Attanasio, C. Meghir, and M. Vera-Hernández, "Baseline report on the evaluation of familias en acción," 04 2004.

[16] O. P. Attanasio, C. Meghir, and A. Santiago, "Education choices in mexico: Using a structural model and a randomized experiment to evaluate progresa," *The Review of Economic Studies*, vol. 79, no. 1, pp. 37–66, 2012.

[17] J. D. Angrist and J.-S. Pischke, "The credibility revolution in empirical economics: How better research design is taking the con out of econometrics," *Journal of Economic perspectives*, vol. 24, no. 2, pp. 3–30, 2010.

[18] A. Armand and P. Carneiro, "Impact evaluation of the conditional cash transfer program for secondary school attendance in macedonia," Tech. Rep. 69, International Initiative for Impact Evaluation, 2018.

[19] M. Li and J. L. Tobias, "Bayesian inference in a correlated random coefficients model: Modeling causal effect heterogeneity with an application to heterogeneous returns to schooling," *Journal of econometrics*, vol. 162, no. 2, pp. 345–361, 2011.

[20] J. M. Wooldridge, "Further results on instrumental variables estimation of average treatment effects in the correlated random coefficient model," *Economics letters*, vol. 79, no. 2, pp. 185–191, 2003.

[21] A. Ertefaie, M. Asgharian, and D. A. Stephens, "Variable selection in causal inference using a simultaneous penalization method," *Journal of Causal Inference*, vol. 6, no. 1, p. 20170010, 2018.

[22] A. Belloni, V. Chernozhukov, I. Fern ndez Val, and C. Hansen, "Program evaluation and causal inference with high-dimensional data," *Econometrica*, vol. 85, no. 1, pp. 233–298, 2017.

[23] A. Belloni, V. Chernozhukov, and L. Wang, "Square-root lasso: pivotal recovery of sparse signals via conic programming," *Biometrika*, vol. 98, no. 4, pp. 791–806, 2011.

[24] R. Tibshirani, "Regression shrinkage and selection via the lasso: a retrospective: Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B, Statistical methodology*, vol. 73, no. 3, pp. 273–282, 2011.

[25] R. Rios, R. J. Miller, N. Manral, T. Sharir, A. J. Einstein, M. B. Fish, T. D. Ruddy, P. A. Kaufmann, A. J. Sinusas, E. J. Miller, T. M. Bateman, S. Dorbala, M. Di Carli, S. D. Van Kriekinge, P. B. Kavanagh, T. Parekh, J. X. Liang, D. Dey, D. S. Berman, and P. J. Slomka, "Handling missing values in machine learning to predict patient-specific risk of adverse cardiac events: Insights from refine spect registry," *Computers in biology and medicine*, vol. 145, pp. 105449–105449, 2022.

[26] M. Fernando, F. Cèsar, N. David, and H. José, "Missing the missing values: The ugly duckling of fairness in machine learning," *International journal of intelligent systems*, vol. 36, no. 7, pp. 3217–3258, 2021.

[27] K. Mohan and J. Pearl, "Graphical models for processing missing data," *Journal of the American Statistical Association*, vol. 116, no. 534, pp. 1023–1037, 2021.

[28] D. Rubin, "Multiple imputations in sample surveys-a phenomenological bayesian approach to nonresponse," *Proceedings of the survey research methods section of the American Statistical Association*, vol. 1, no. 1, pp. 20–34, 1978.

[29] V. Chernozhukov, D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins, "Double/debiased machine learning for treatment and structural parameters," *The econometrics journal*, vol. 21, no. 1, pp. C1–C68, 2018.

[30] I. N. d. E. y. C. d. C. R. INEC, "Registro encuesta nacional de hogares." `http://sistemas.inec.cr/pad5/index.php/catalog/REGENAHO`, 2022. Accessed: 2022-08-28.

[31] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.