STOCKHOLM SCHOOL OF ECONOMICS
Department of Economics
5350 Master's thesis in economics
Academic year 2019-2020

Thank God, It's a Boy?! The Heterogeneous Effects of Children's Gender on Domestic Violence

An Application of the Causal Forest Algorithm

Carolin Thomas (41452)

Abstract This thesis aims to quantify heterogeneity in treatment effects of the gender of the first child on the mother's probability of experiencing domestic violence. The identification strategy takes into account the potential endogeneity of gender preferences on the children's gender, as well as potential channels of a child's gender on domestic violence. I use a large-scale data set from Colombia with close to 50,000 observations and rely on the causal forest algorithm introduced in Wager and Athey (2018) to quantify heterogeneous treatment effects. I identify mainly two different subgroups of women for which the size of the treatment effect differs around six percentage points. Women who are less likely to experience domestic violence when having a first-born son compared to a daughter are on average younger, have a larger age gap to their partner and are more likely to live in rural areas. Women who are more likely to experience domestic violence when having a first-born son compared to a daughter are on average older, are living in cities and more likely part of the upper social classes. The robustness checks provide evidence on, among others, reporting, attrition and sample bias.
Keywords: Domestic Abuse, Preferences for Sons, Colombia, Causal Forest, Heterogeneous Treatment Effects

JEL: C55, J12, J13, N36

Supervisor: Andreea Enache
Date submitted: May 17, 2020
Date examined: May 27, 2020
Discussant: Camille Habé
Examiner: Kelly Ragan

Contents

1 Introduction

Around 30% of all women worldwide are currently experiencing domestic violence or have experienced domestic violence through the hands of their husband/partner. The exact percentage differs by region and ranges between 23% in high-income European countries and 46% in African countries (Garcia-Moreno et al., 2013).[1] In 2017, around 30,000 women were intentionally murdered by their husbands/partners, which results in an average of 82 women on every day of the year (UNODC Vienna, 2019). However, Economics of violence against women and specifically the field of domestic violence is still one of the lesser known and explored research areas of Economics. Consequently, research in terms of explanatory models and rigidly conducted analyses is scarce.

Domestic violence is challenging to study due to its nature. In general, it refers to all acts of violence committed in the domestic environment, which makes partners, children, the own family and the family in law potential victims and offenders. The violence can be split in several forms, such as physical violence, which itself is usually divided in different categories, depending on the severity and nature of the violence, such as beating, strangling, burning or raping of the partner and emotional violence, such as controlling behavior. The additional perversity of the crimes is the lack of incentives for and the surrounding fear of the survivors to report the crimes, intensified by societal pressure and cultural norms. Contrary to crimes that are committed by strangers, survivors know the offenders and are often emotionally attached to them. Besides the emotional struggle, reporting the crime and a potential subsequent conviction will not only affect the life of the offender negatively, but extremely likely also the life of the survivor. Examples are stronger financial constraints due to a lower family income, threats and aggressions by the family side of the offender and sole responsibility of the common children, if there are any. Additional problems surround the legal nature of the violence, which differs between countries and sometimes makes it impossible to report the experienced violence. In 2014, only 35 out of 160 countries legally considered all forms of domestic violence listed above a crime (Oecd, 2016). In comparison, all types of sexual harassment were outlawed in 41 out of 160 countries in the same year (Oecd, 2016). Taking all of these factors together results in less than 40% of women and girls affected by violence reporting the crimes or seeking help (United Nations, 2020). The direct and indirect costs of domestic violence, such as hospital bills and sick hours due to obtained injuries are high. An estimate of the global costs of violence against women amounts to $1.5 trillion (United Nations, 2020). Given the numbers mentioned above, the need to improve the current situation for survivors of domestic violence is apparent. However, there is currently not enough knowledge on causal determinants for domestic violence behavior and efficient solutions to support survivors.

This thesis studies one potential risk factor of domestic violence in detail and makes contributions to several fields in Economics.[2] The exact research question is the following: Does a father's unfulfilled desire for sons affect his violent behavior towards his female wife/partner and how does the effect depend on the characteristics of the woman? I attempt to answer this question by studying the impact of the gender of the first child on the reported domestic violence committed by the husband/partner, using a large-scale data set from Colombia containing around 50,000 women and Machine Learning methods. Colombia is among the countries with the highest rates of domestic violence in the world. One estimate of costs of domestic violence in Colombia amounts to around 7.14 billion pesos or 3.2% of GDP in 2003, accounting only for lost labor earnings (Ribero and Sánchez, 2005). Throughout this thesis, I will refer to women whose first-born child was a son as "treated" and to women whose first-born child was a daughter as "untreated". Accordingly, the term "treatment" refers to a first-born son.

[1]Comparable numbers for men are more difficult to obtain.

[2]This thesis was mainly written during my stay in the Machine Learning and Causal Inference lab of Susan Athey at the Stanford Graduate School of Business, lasting throughout the year 2019. I use the R package 'grf' for the main analysis, which is maintained at Stanford.

The research question of this thesis has, to the best of my knowledge, not yet been the focus of any study of risk factors of domestic violence. Most previous attempts that are similar to mine control for the overall gender composition of the children when working on domestic violence and do not take into account the potential endogeneity of gender given gender preferences, nor potential channels of children's gender on domestic violence. I argue for the importance to include both in Section 3 and I propose an identification strategy that allows to do so. Furthermore, I use non-standard methods to evaluate the impact of children's gender, not only on the average women in the sample, but on specific subgroups of women, whereby the subgroups are defined over observable characteristics. The data-driven approach allows to determine heterogeneous treatment effects and relies on the exogeneity of the gender of the first child, for which evidence will be provided in Section 4.3. The identification of smaller subsamples of women for which the treatment has different effects gets more relevant, the lower we expect the number of affected women to be. I discuss the need for new methods to detect heterogeneous treatment effects in Section 5.

The causal forest method, which is used in this thesis and has been introduced in Wager and Athey (2018), is a forest-based Machine Learning method and allows to identify causal treatment effects. Common tree and forest Machine Learning methods aim to maximize the difference in outcomes to identify similar subgroups of individuals in terms of the outcome and to identify variables related to the difference in outcomes. The causal forest algorithm has been adapted to maximize the difference in average treatment effects of certain subgroups and, to some degree, allows to identify variables related to the difference in treatment effects. Section 5.1 starts with Machine Learning methods in general, Section 5.2 introduces the main principles of forests and in specific, of the causal forest and Section 5.3 and Appendix Section Additional Explanations to the Causal Forest and Regression Forest Algorithm explain the estimation procedure of the causal forest. Here, the causal forest is used to identify certain subgroups of women whose partners have reacted differently to the gender of the first child in terms of violent behavior. The need for Machine Learning methods in this thesis is based on several reasons:

First, I am interested in heterogeneous treatment effects, mainly because I do not expect the average woman to be extensively affected by the risk factor I am studying. Instead of manually trying different specifications of subsamples, which is highly problematic for the identification of valid treatment effects as demonstrated in Section 5, the causal forest is a data-driven approach and results in valid estimates of the treatment effect.

Second, common methods such as the linear probability model (LPM), logit and probit, for which I provide results in Section 6 are very likely to not provide valid estimates of the treatment effect in this setting. This is due to the large difference in propensity scores for treatment and control group, estimated in this thesis using a regression forest. This difference is caused by the sample selection necessary for the identification strategy. Section 4.4 contains evidence on the difference in the probability of having experienced domestic violence for women who are (not) included in the sample and for women in the sample with and without a first-born son, as well as additional analyses on the differences in characteristics between these groups and on potential drivers of the difference in propensity scores. I show that important differences in observable characteristics between the control and treatment group are statistically significant, as is the mean of the propensity score for the two groups. The causal forest estimates the treatment effect while taking into account the propensity score and therefore results in estimates with a higher validity for the average treatment effect, as well as for the heterogeneous treatment effects.

I find that there are mainly two different subgroups of women for which the treatment effect differs around six percentage points and which will be discussed in detail in Section 6.2. Contrary to common methods, the causal forest does not result in a single measure to evaluate the output. Instead, I present several aspects of the output which allow to get a clearer picture of the women in the identified subgroups. Women who are less likely to experience domestic violence when having a first-born son compared to a daughter are on average younger, have a larger age gap to their partner and are more likely to live in rural areas. Women who are apparently more likely to experience domestic violence when having a first-born son compared to a daughter are on average older, are living in cities and are more likely part of the upper social classes. All of these results have to be considered with caution, which I demonstrate in Section 7. I provide additional results of the causal forest based on different definitions of the treatment and the subgroups. Furthermore, I present evidence on data concerns such as sample, reporting and attrition bias. I attempt to determine, among others, during which period of their life women are most likely to experience domestic violence and if the number of children or the number of sons affect the probability of having ever experienced domestic violence. To provide information on potential attrition bias in the sample, I predict the value of one of the previously identified main drivers of heterogeneity in treatment effects for a subset of women with missing values in this variable using a regression forest and rerun the causal forest algorithm.

The thesis is structured in the following way: Section 2 provides an overview of the current literature on risk factors for experiencing domestic violence and the existence of preferences for sons, as well as previous attempts to combine the two literatures. In Section 3 I present the identification strategy and Section 4 contains information on the data set and the exogeneity of the treatment. Section 5 introduces the causal forest which is used to estimate heterogeneous treatment effects and Section 6 contains the main results of common estimation methods and of the causal forest. In Section 7, I conduct several robustness checks to quantify major data concerns, such as potential reporting, sampling and attrition bias. Section 8 discusses the presented results and potential policy implications in detail and concludes.

2 Literature Overview

2.1 Risk Factors for Experiencing Domestic Violence

It is difficult to determine the reasons why some women experience domestic violence. Risk factors can work as indicators to identify women who have a higher risk of experiencing domestic abuse due to the characteristics of their partners or husbands, characteristics of their relationships and their own characteristics.

Personal Characteristics Looking specifically at Latin American countries, Flake and Forste (2006) find that living together with the husband/partner, a higher degree of female autonomy in decision making and the husband consuming alcohol increase the probability of experiencing domestic violence in Haiti, Colombia, Peru, Nicaragua and the Dominican Republic. Only in Colombia and Haiti does a woman with a lower level of education than her husband/partner have a higher probability of experiencing domestic abuse.

González-Brenes (2003) investigates the risk factors for domestic violence in Tanzania and finds that young women who are still in the family planning process are especially vulnerable, as well as women with small female family networks. She finds no connection between household wealth and the prevalence of domestic violence.

Oestby et al. (2019) study domestic violence in the aftermath of the war in Peru and find that among all married couples, women are most likely to experience domestic violence in the early years of their marriage and if their husband consumes alcohol. Furthermore, domestic violence is more prevalent among middle-income households. The woman's education does not impact the probability of experiencing domestic violence, but the education of the husband does inversely.

The review of Semahegn and Mengistie (2015) on risk factors in Ethiopia shows that among others, the alcohol consumption of the husband, drugs, a family history of violence, being married and living in a rural area makes women more likely to experience domestic violence. Contrary to that, González-Brenes (2004) finds that in East Africa, women living in cities, in polygamous relationships and women who have been married for a longer time are at a higher risk of experiencing domestic violence. Similar to González-Brenes (2003), women with an extended network of female family and friends are less likely to be abused by the husband/partner.

Vyas and Watts (2009) review 24 papers on the impact of women's empowerment on domestic violence in low-income countries and find that having a higher education and a higher level of household assets reduce the probability of domestic violence. The impact of working status was ambiguous across the papers.

Income Due to the endogeneity of most labour market outcomes it is difficult to determine causal effects. In theory, the effect of female empowerment in the form of paid work on domestic violence can go in both directions. A higher relative female income improves the female status in the relationship and decreases the opportunity costs of leaving an abusive relationship, which should reduce domestic violence. On the opposite, an empowered women can be perceived as threat to the typical role distribution in a relationship by her husband/partner who acts more violently to restore the patriarchic balance (Vyas and Watts, 2009).

Angelucci (2008) relies on a randomized control trial (RCT) in Mexico that entitled women to a conditional cash transfer (CCT) and simultaneously reduced the alcohol abuse of the husband/partner by on average 15%. She finds that through the channel of alcohol, the average level of domestic violence decreased. However, women who were married to an older, less educated husband, experienced an increase in domestic violence, which can be explained by the perceived threat of traditional roles in a relationship. The author advocates that this harmful effect has to be taken into account when administering cash transfers to households. Bobonis et al. (2013) investigate the same setup in Mexico and find that women who were randomly attributed to receive the money were five to seven percentage points less likely to experience physical domestic violence, but three to five percentage points more likely to be survivors of emotional violence.

Hidrobo and Fernald (2013) use a RCT in Ecuador that handed out unconditional cash transfers (UCT) to women to determine the effect of a pure increase in income on domestic violence. The authors find that the effect depends largely on the relative education of wife and husband. If the wife had less than primary education and the husband had received less schooling than her, the cash transfer resulted in an increase in emotional violence. On the other hand, women with more than six years of schooling experienced a decrease in emotional violence following the unconditional cash transfers.

Chin (2012) uses exogenous variation of rain shocks and studies its impact on the labour market in rural India. The author estimates that women who were receiving a salary from their work were less likely to experience domestic violence. In this study, the reduction in violence can mostly be explained by less exposure to violent behavior, since the women spent more time working outside of the home, rather than through an increase of the women's status. Heath (2014) looks into the relationship of working status and experienced domestic violence and finds that working women who were married early and received little education have a higher probability of being abused by their husband/partner. However, the author ignores the potential reverse causality of domestic violence on the working status.

Aizer (2010) investigates the exogenous variation in the gender wage gap and finds that an increase in the relative wage in industries dominated by women lead to a decline in experienced domestic violence in California between 1990 and 2003, measured by physical injuries registered in emergency rooms of hospitals. Using a similar measure of domestic violence, Iregui-Bohórquez et al. (2019) study the effect of labor income in rural Colombia between 2009 and 2013 on a municipality level. According to the authors, an increase in income reduces domestic violence in most work sectors, such as agriculture. Contrary, women working in services experience an increase in domestic violence.

Education Erten and Keskin (2018) use a regression discontinuity design based on a school reform in Turkey in 1997. The authors find that an increase in female education lead to higher reports of emotional violence, but reports of physical violence were unaffected. Based on the results, it is unclear what share of the effect can be attributed to a difference in the reporting behavior due to a higher education instead of an actual change in experienced domestic violence.

2.2 Parents' Preference for Sons

Preferences for sons can have different reasons and it is important to take into account that not all of the explanations necessarily have to be purely gender-biased. Instead, it can be rational to prefer sons in a given setting as demonstrated by the reasons listed below:

Roles Fathers are expected to act as role models for their sons, which explains why they spend more resources on and time with them.

Support Fathers might expect their sons to take over or support them doing the "male tasks" in the household (Godoy et al., 2006).

Costs The costs of bringing up girls and boys might be different. If costs for girls are higher and lead to the same benefit, parents prefer having sons.

Development The development of boys and girls might be different in the early stages of childhood. This requires families to spend more time with their sons or to spend more money on their son's health.

Future Sons and girls might have different prospects for the future. Since sons are more likely to work and to have a professional career in many countries, spending money on their education yields a higher payoff.

In the US, Dahl and Moretti (2004) and Raley and Bianchi (2006) find that parents with girls are less likely to remain married and divorced fathers are more likely to have custody for their sons compared to their daughters. Women who are expecting a boy and are unmarried at conception, are more likely to have a shotgun marriage before the delivery, whereas women with only girls have a lower probability of ever being married. In families where the first two children have the same sex, families with a first-born girl are more likely to have a third child, compared to families with a first-born boy (Dahl and Moretti, 2004).

When asked explicitly if they would rather wish for a son or a daughter, women only show a slight preference for daughters, whereas men are twice as likely to prefer sons (Dahl and Moretti, 2004; Bharadwaj et al., 2014). On social media, parents mention their sons more often than their daughters and posts with sons get more likes on average (Sivak and Smirnov, 2019).

In general, fathers are more involved in the upbringing of their sons. According to Mammen (2011), fathers in the US with at least one son spend more time with their children and take the additional time from their leisure. The additional hours spent with the children also benefits the girls with at least one brother who get more time with their father.

Among Asian immigrants in the US, only mothers are more likely to spend additional time with their newborn sons (Kaushal and Muchomba, 2018). Ichino et al. (2014) shows that women in the US, the UK, Italy and Sweden are more likely to work the common amount of hours per week and even take on additional hours after having had a first-born girl compared to a first-born son.

The literature on preferences for sons in Colombia is scarce. Dahl and Moretti (2004) find that among parents with a first-born girl, the probability is statistically significantly higher to be divorced compared to having had a first-born boy.

2.3 The Effect of Children's Gender Composition on Domestic Violence

There is only a limited literature on the effect of the composition of the children's gender on domestic violence experienced by the mother through the father of the children. The majority of the papers that I am aware of do not control for the potential endogeneity of a child's gender nor for potential channels of the gender on domestic violence. This thesis contributes one potential way on how to deal with these two issues. Finnoff (2012) investigates domestic violence in Rwanda during the years after the genocide and finds that women with at least one son who was living at home did not report a different rate of emotional or sexual domestic violence, but did report a higher rate of physical domestic violence compared to women with only girls. Women who have lost a son are also more likely to experience domestic violence. González-Brenes (2004) looks at the impact of household composition on domestic violence in East Africa. She finds that a teenage son living at home reduced the probability of having experienced domestic violence during the last year, which could also be due to the son intervening during domestic abuse. The death of a son increased domestic violence in a household, whereas the total number of sons or the ratio of sons to girls did not have an impact.

Somville (2020) is, to the best of my knowledge, the paper that is closest to this thesis in terms of the research question and identification strategy and has only been recently published as a working paper. The author studies whether fathers with a first-born daughter instead of a first-born son are less likely to be violent against their female wife/partner and finds that reported domestic violence decreases in households with a first-born girl. He relies on a data set of 310,000 couples from 18 different African countries that were surveyed between 2006 and 2017. Similar to this thesis, he relies on the assumption that the gender of the first child is exogenous. However, the assumed underlying mechanisms for the treatment effect differ. This thesis focuses on potential preferences for sons as the reason for an increase or decreases in domestic violence, whereas Somville (2020) studies whether having a daughter leads to an emotional change of the father's perspective towards domestic violence. This difference also affects the identification strategy, since Somville (2020) includes all couples with at least one child and I impose further and more severe restrictions that are tailored towards the channel I study. Furthermore, the respective goal of the analysis and therefore, the methods used are different. Somville (2020) uses an OLS specification to quantify the average treatment effect, whereas I am interested in potential heterogeneous treatment effects and rely on Machine Learning methods for the identification.

3 Identification Strategy

Why is gender endogenous?

Without further thought, one generally expects the gender of a child to be randomly distributed and therefore to be exogenous to the characteristics of the parents. This is currently assumed in most of the papers trying to quantify the effect of children's gender on domestic violence. However, there are several reasons why this is most likely not fulfilled in reality, which emphasizes the need to take potential endogeneity into account:

First, some literature suggests that wealthier and better educated parents are more likely to have sons compared to poorer, less educated people. This phenomenon is called Trivers-Willard hypothesis (Trivers and Willard, 1973).

Due to the fact that this is very difficult to quantify or control for, the hypothesis is commonly ignored in the Economics literature.

Second, the existence of preferences for sons threatens or even compromises the assumed exogeneity of gender, whereby the extent of the preferences influences the validity of this assumption. The most extreme form can be observed in Asian countries and among immigrants of Asian descent in Western countries (such as the US, see Abrevaya (2009)). Among these groups, preferences for sons lead to sex-selective abortions after the gender was determined during the pregnancy and to negligence of female children after birth, which results in a higher rate of infant mortality among girls in these countries or ethnic groups (e.g. Barcellos et al. (2014) investigates the effect of preferences for sons in India). Therefore, the probability of having a living son can depend on the preferences for sons, which could itself be conditional on characteristics of the family. In Western countries, preferences for sons are prevalent even if sex-selective abortions are extremely rare. Instead, the preferences can be seen, among others, through a phenomenon called the "son-stopping rule". This rule describes the fact that the probability to try for an additional child increases when the children before were all girls compared to all boys. On average, this leads to girls growing up in families with more children (Barcellos et al., 2014). Since the larger family size can also influence financial and educational aspects of the families and preferences for sons are potentially more likely to exist in families with specific *a priori* characteristics, girls and boys can on average grow up in families that are not only different in the number of children. In this case, the gender of a child can no longer be seen as exogenous.

There are several ways proposed in the literature on how to deal with the endogeneity of children's gender. It is important to note that not all papers interested in the effects of gender composition of children take the potential endogeneity into account.

The most common approach relies on the following assumption: The gender of the first child can largely be seen as exogenous to family characteristics (ignoring the Trivers-Willard hypothesis) and family size (Bharadwaj et al., 2014). Ichino et al. (2014) use all first-born children younger than 15 to determine differences in work patterns between women with a first-born son or girl. Choi and Hwang (2015) use the gender of first-born children to determine whether preferences for sons still exist in Korea.

Depending on the exact research question, it is useful to set the additional restriction that the gender of the first-born child is only exogenous as long as the parents did not have time to react to its gender (Barcellos et al., 2014). Kaushal and Muchomba (2018) restrict the sample to families whose child is younger than two and explores whether Asian mothers in the US spend more times with sons compared to daughters during the early childhood. Since having a first-born girl increases the likelihood of having a second child, the time spend with first-born girls is on average lower when not using the age restriction c.p. Novella (2019) restricts the sample to children less than one year old to determine whether allocation of food during the early childhood is affected by the parents' gender preference. Lundberg et al. (2007) use the gender of the first-born child and examine the behavior of the father during the first year after the birth. Alternative strategies are used in Mammen (2011) who divides families in two groups, depending on whether they have at least one son and compares the time fathers spent with their children. Shafer and Malhotra (2011) rely on a panel-design approach by using parents' answers on gender-related statements before and after the birth of their first-born child.

Assumptions

The gender of a child should not have an effect on domestic violence in the absence of gender preferences. However, when gender preferences exist, frustration over not having a child of the preferred gender, assumed to be male based on the evidence presented in Section 2, can lead to an increase in domestic violence committed by the father of the child.

Assumptions:

1. The domestic violence experienced due to the unfulfilled desire for a son stops as soon as a son is born and does not depend on the number of sons born, nor the number of children born in total. After having had a son, the decision on having further children is only based on the desired number of total children, regardless of the gender.
2. The gender of the first child is exogenous to family characteristics (ignoring the Trivers-Willard hypothesis) and there is no practice of sex-selective abortions. The gender of the first child should not be influenced by the preferences for a son by the father/expecting parents in the absence of sex-selective abortions.
3. Family size does not have an impact on domestic violence or vice versa. Furthermore, the duration of exposure (measured in years) does not impact domestic violence. Women who have had more children are more likely to have been exposed longer to their husband/partner. This assumption is supported by findings that show that domestic violence in Colombia is most prevalent during the first two years of marriage (Kishor and Kiersten, 2004).
4. The reporting behavior of domestic violence is exogenous among all women.

Sample Selection

In order to control for the extent of the preferences for sons, which are generally unobserved, I only include women in the sample who have had at least one living son as an indicator of existing preferences for sons. Therefore, the husbands/partners of the women in the sample are likely to have similar preferences for sons and less likely to have preferences for girls. Given the assumptions above, the expected channel of the treatment in this thesis is the following: Women whose first-born child was a son have never experienced domestic violence due to an unfulfilled desire for sons by their husband/partner. All other women in the sample have had at least one girl before they gave birth to a son. During this time the desire for a son was unfulfilled and these women subsequently experienced a change in the domestic violence committed by their husband/partner. Since all women in the sample have at least one son, but the birth order by gender is different, I expect women whose first-born child was a girl to report a different level of experienced domestic violence on average.

Only including women with at least one son has several effects on the sample and makes the distribution of certain characteristics in the sample not random as demonstrated in Subsection 4.3.

First, I exclude women with only girls. These are women whose husband/partner has (still) an unfulfilled desire for sons or whose husband/partner has no preferences regarding the gender of the children or preferences for girls. Excluding the two last-mentioned has the desired effect on the sample. However, women who never gave birth to a son and have completed the family planning process have probably experienced more domestic violence compared to women included in the sample. Similarly, women with a first-born girl and subject to the same preferences who are still in the family planning stage and are trying to get a son are excluded. This leads to an artificial restriction of the sample size of the control group.

Second, I oversample families who only had one child and are still in the family planning stage in the treatment group and I undersample these women in the selected sample compared to all women in the entire data set. Simply oversampling families with only one child who turned out to be a boy is not critical, since these are the ones that are very likely to have strong preferences for sons as predicted by the son-stopping rule and prefer to have the least amount of children possible, conditional on having a son. Women who are still trying to get more children have already fulfilled their husband's/partner's desire for a son and are therefore less likely to experience domestic violence compared to women whose first-born child was a girl. This holds true if exposure to the husband/partner does not have an impact on experienced domestic violence. Still being in the family planning stage also impacts other characteristics, such as the age of the respondent. It is not possible to restrict the sample to families with more than one child to control for the birth order, since this will exclude those with the strongest preferences for sons who followed the son-stopping rule. Due to the sample restrictions, the probability that fathers react differently to a first-born son is reduced.

Potential channels of gender on domestic violence

As demonstrated in Subsection 2.2, the gender of the child can influence the behavior of the mother and as shown in Subsection 2.1, these behaviors can be risk factors to experience domestic violence. For example, if a mother of a newborn son is less likely to start work again after birth, the higher budget constraint the family is facing can lead to an increase in tension and domestic violence. However, the current literature on risk factors for domestic violence rarely controls for these potential channels or bad controls of the child's gender on domestic violence. In order to do so, I exclude all variables related to the relationship and current working status, as well as some characteristics of the household, the number of children and current pregnancies.

An additional argument to exclude these variables is provided by the Machine Learning methods used in this thesis (see Athey et al., 2018), which require to limit the set of covariates to pre-treatment variables. This step verifies that the values of the included variables is determined prior to the introduction and therefore independent of the treatment, which is fulfilled when excluding the following set of covariates:

Relationship The child's gender can influence the relationship status of the mother such as whether she is married, separated, divorced or living together with the father and her number of relationships/unions/marriages (Dahl and Moretti, 2004), which can itself influence the probability of domestic violence (Semahegn and Mengistie, 2015).

Work The child's gender can influence whether the mother is currently working (Ichino et al., 2014; Kaushal and Muchomba, 2018) and the working status can impact the probability of experiencing domestic violence (Chin, 2012; Heath, 2014).

Household As mentioned in Subsection 3, the household size, including the number of children, can depend on the gender of the first child, as well as female autonomy, which can make the women more vulnerable to domestic abuse (Flake and Forste, 2006).

Pregnancies A current pregnancy and terminations in the past can be influenced by the gender of the first child (see the son-stopping rule (Barcellos et al., 2014)), which can also affect the violent behavior of the husband/partner.

4 Data

4.1 Overview

I am using three data sets on Colombia from the years 2004 & 2005,[3] 2009 & 2010[4] and 2015 & 2016,[5] which are part of the internationally well-established Demographics and Health Survey (DHS). The DHS is one of the very limited sources for data on domestic violence and is cited in the majority of papers in this field. The organization behind the DHS provides a standardized questionnaire, which can be used by local agencies or the government in developing countries. The surveys in Colombia are conducted in a five-year cycle by Profamilia, an NGO, and target a nationally representative sample of women between the age of 13 and 49.

The sample, which I will refer to as the main sample throughout the remainder of the thesis, includes all women who have ever been in a union or marriage, never had a multiple birth, such as twins or triplets, and have at least one living son, as well as non-missing outcomes and non-missing values for the main covariates listed in Table 2.[6] Table 19 in the Appendix shows how the sample composition is affected by each of the sample constraints. The outcomes are among the standard outcomes in the literature and follow the guidelines developed by the WHO (see Garcia-Moreno et al. (2005)) on how to ask questions on domestic violence and how to categorize experienced domestic violence.[7] The three outcomes are indices on less severe violence, severe violence and sexual violence and take the value 1 if the woman has ever experienced at least one of the items listed below by her husband/partner in a given period of time:

Less severe violence: pushed, shook or had something thrown towards her; slapped; punched with fist or hit by something harmful

Severe violence: kicked or dragged; strangled or burnt; threatened or attacked with a weapon

Sexual violence: physically forced into unwanted sex

The setup of the interview does not allow to identify explicitly whether the father of the child(ren) abused the respondent. This implies that I cannot be certain that the abuser and the father of the children are the same person. However, since the majority of women are married or in a relationship with the father of the child(ren) and the outcomes indicate whether the woman experienced violence by her husband/partner, the violence is most likely committed the father of the child(ren).

4.2 Summary Statistics

Outcome and treatment

Table 1 shows the summary statistics for the outcomes and the treatment and the information available on the timing of experienced domestic violence for the main sample as defined in Subsection 4.1. The main sample contains 46,922 women out of which 68 percent have a first-born son. The reason that the treatment and control group differ in size is due to the sample composition. Out of all women with one living child, approximately half

[3] Ministerio de Salud y Protección Social y Profamilia [Producers] (2005)
[4] Ministerio de Salud y Protección Social y Profamilia [Producers] (2011)
[5] Ministerio de Salud y Protección Social y Profamilia [Producers] (2017)
[6] The last constraint is due to the Machine Learning methods used in the thesis which cannot deal with missings. In Section 7.3 I estimate the missing values of one main covariate based on the values of the remaining non-missing covariates using a regression forest.
[7] E.g. instead of asking whether the women is “a survivor of domestic violence”, she is asked whether she has ever experienced specific types of violence by her husband/partner. Furthermore, confidentiality has to be guaranteed and the questions are only asked in the absence of any person other than the respondent and the interviewer in the same room.

are included in the main sample since they have a boy. Out of all women with two living children, approximately 75 percent are included in the main sample since they have at least one son and this goes on until the 14th child which is the maximum of living children any woman has in the main sample. Therefore, I oversample young women with only few children in the treatment group and I undersample women with only few children in the main sample compared to the entire sample. In Subsection 4.3 I will show the differences in observable characteristics between treatment and control group and in Subsection 4.4 I will discuss in further detail how the main sample used in the paper is different from the entire sample of women questioned by the DHS.

In the main sample, around a third of all women have ever experienced domestic violence by their husband/partner throughout their life.[8] The majority of reported domestic violence is categorized as less severe domestic violence, which is experienced by 28 percent of all women in the sample. Ten percent of all women have experienced severe violence in their relationship at least at one point during their life and seven percent were raped by their husband/partner. Most incidences of less severe and severe violence is reported to have happened before the last year prior to the interview.

Figure 1 shows the pairwise correlation between the outcomes and the treatment. The correlation between the treatment and outcomes is low and negative, especially for the outcomes that happened during the last year. The strongest correlations can be observed between the same outcomes that were ever experienced or during the last year, and the correlations are slightly lower for the outcomes experienced before the last year. The correlation between less severe violence and severe violence are the strongest among all combinations of the three outcomes, showing that in general, women who experience less severe violence are also more likely to experience severe violence, which indicates an escalation of the level of domestic violence.

Figures 2 shows the distribution of the different forms of domestic violence compared to the treatment on a departmental level for all continental departments in Colombia. The treatment differs between 0.62 and 0.72 across the departments. Since the value of the treatment is generally expected to be around 0.5 for all departments, this shows interesting features of the departments due to the sample composition. The higher the value of the treatment in a department, the higher has to be ratio of women who are young and potentially still in the family planning process and only have a few children, most likely only a son. Since the sample is representative, the departments with a higher treatment value are likely places with more young people in general. There seems to be no obvious relation between the treatment and domestic violence patterns on the department level.

Main Covariates

The main covariates included in the analysis are summarized in Table 2 and have only non-missing values as defined in the sample composition in Section 4.1. On average, the husband/partner of the women questioned is slightly more than four years older and the woman herself is around 35 years old. Most of the women work in the service sector and went to school for an average of less than nine years.[9] 35% of all women in the sample grew up in a household in which the woman's mother experienced domestic violence by the woman's father/step-father.

[8]When drawing a comparison to the level of domestic violence experienced in Colombia, it is important to take into account that only women up to 49 years are interviewed by the DHS. Overall experienced domestic violence is likely higher among all women in Colombia.

[9]The variables "works in .. " indicate which field of profession the woman has chosen and does not imply that she is currently working in this field. The left-out category indicates that she hasn't chosen a field. The variables are mutually exclusive.

	q5	mean	median	q95	N
Has experienced at least one of the three outcomes					
ever	0.00	0.30	0.00	1.00	46922
before last year	0.00	0.23	0.00	1.00	46922
during last year	0.00	0.18	0.00	1.00	46922
Outcome: Less severe violence					
ever	0.00	0.28	0.00	1.00	46922
before last year	0.00	0.20	0.00	1.00	46922
during last year	0.00	0.17	0.00	1.00	46922
Outcome: Severe violence					
ever	0.00	0.10	0.00	1.00	46922
before last year	0.00	0.08	0.00	1.00	46922
during last year	0.00	0.06	0.00	1.00	46922
Outcome: Sexual violence					
ever	0.00	0.07	0.00	1.00	46922
before last year	0.00	0.04	0.00	0.00	46922
during last year	0.00	0.04	0.00	0.00	46922
Treatment					
Has at least one living son	1.00	1.00	1.00	1.00	46922
Her first child is male	0.00	0.68	1.00	1.00	46922

Table 1: Summary statistics for outcomes and treatment

The percentiles and the mean are weighted with the domestic violence sample weight provided by DHS. The sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. Less severe violence indicates whether the woman has ever experienced at least one of the following items: pushed, shook or had something thrown; slapped; punched with fist or hit by something harmful. Severe violence indicates whether the woman has ever experienced at least one of the following items: kicked or dragged; strangled or burnt; threatened or attacked with weapon. Sexual violence indicates whether the woman was ever physically forced into unwanted sex. All outcomes only refer to actions committed by the husband or partner.

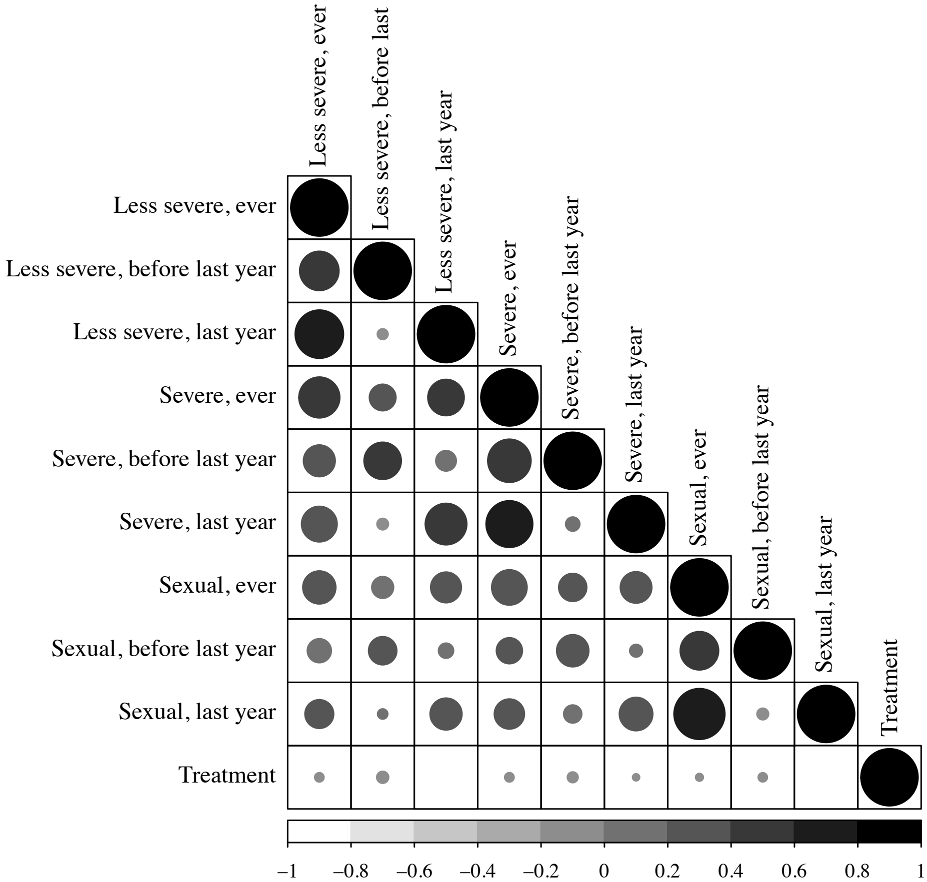

Figure 1: Correlation between outcomes and treatment

The figure shows the pairwise correlation for all pairs for which the p-value of the correlation is higher than 0.05. The size of the circles indicates the absolute value of the correlation. The sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. Less severe violence indicates whether the woman has experienced at least one of the following items: pushed, shook or had something thrown; slapped; punched with fist or hit by something harmful. Severe violence indicates whether the woman has experienced at least one of the following items: kicked or dragged; strangled or burnt; threatened or attacked with weapon. Sexual violence indicates whether the woman was physically forced into unwanted sex. All outcomes only refer to actions committed by the husband or partner.

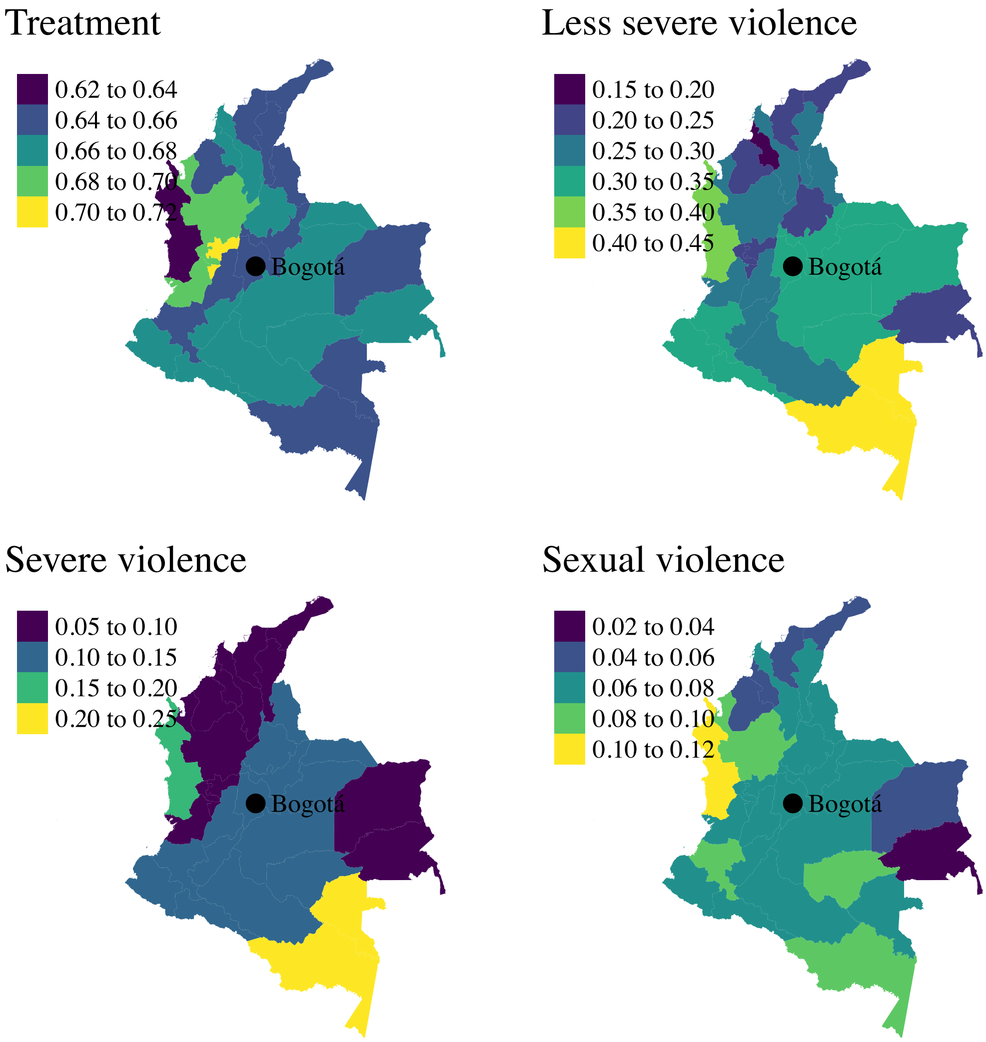

Figure 2: Map of percentage of women who have ever experienced domestic violence

The maps show the weighted mean of the indicated variable on a department level, weighted with the domestic violence sample weight provided by DHS and include all continental departments of Colombia. The sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. Less severe violence indicates whether the woman has ever experienced at least one of the following items: pushed, shook or had something thrown; slapped; punched with fist or hit by something harmful. Severe violence indicates whether the woman has ever experienced at least one of the following items: kicked or dragged; strangled or burnt; threatened or attacked with weapon. Sexual violence indicates whether the woman was ever physically forced into unwanted sex. All outcomes only refer to actions committed by the husband or partner. The map data was provided by Runfola, D; Seitz, L; Hobbs, L; Panginaban, L; Oberman (2020)

	q5	mean	median	q95	missings	N
Characteristics of the household						
lives in urban environment	0.00	0.73	1.00	1.00	0.00	46922
wealth quintile of household	1.00	2.92	3.00	5.00	0.00	46922
Characteristics of the female respondent						
age difference to partner	-6.00	4.28	3.00	16.00	0.00	46922
age	21.00	35.10	35.00	48.00	0.00	46922
age at first child	15.00	20.35	20.00	29.00	0.00	46922
was minor at first sexual intercourse	0.00	0.71	1.00	1.00	0.00	46922
experienced domestic violence as child	0.00	0.35	0.00	1.00	0.00	46922
works in agriculture	0.00	0.04	0.00	0.00	0.00	46922
works in clerus	0.00	0.07	0.00	1.00	0.00	46922
works in sales	0.00	0.24	0.00	1.00	0.00	46922
works in services	0.00	0.38	0.00	1.00	0.00	46922
works in skilled & manual job	0.00	0.05	0.00	0.00	0.00	46922
works in technical & managerial job	0.00	0.08	0.00	1.00	0.00	46922
works in unskilled & manual job	0.00	0.02	0.00	0.00	0.00	46922
total years of education	1.00	8.27	9.00	16.00	0.00	46922

Table 2: Variables included in the main analysis

The percentiles and the mean are weighted with the domestic violence sample weight provided by DHS. The sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. The variable "experienced domestic violence as child" refers to actions committed by the father against the mother during childhood. "Minor" refers to all women younger than 18 years.

Table 20 in the Appendix shows the summary statistics of additional variables that are excluded in the analysis due to the endogeneity concerns mentioned in Subsection 'Potential channels of gender on domestic violence'. On average, the women are living together with four other family members, have 2.71 children and 18% of the households in the sample have a female head.[10] 95% of all women are either living together with their partner or are married. 27% have ever terminated a pregnancy. At the point of the interview, 54% of the women were currently working and 3% were pregnant. 71% of the women can read and/or write at least one sentence.

4.3 Exogeneity of Treatment

Sex-selective abortions in Colombia

Overall, there is no evidence in Colombia for sex selective abortions or a different rate of infant mortality by gender. Wang et al. (2019) find a constant level of the sex-ratio in Colombia since 1967 at the natural rate of 1.05 boys born for every girl, which is also close to the average around the world. Additionally, no large difference for the sex-ratio for separate age group can be found. If this was the case, it would indicate a preference for sex-selective abortions after the introduction of the medical possibility to detect the gender before birth. The ratio of males to females is lower in Colombia for children and teenagers compared to the worldwide average, whereby the latter is driven by sex-selective abortions in China and India. When taking into account the birth-order of the children

[10]The female household head can either be the respondent herself, a female relative of her or of her husband.

during the years 2009 to 2013, Colombia has a constant ratio of around 107, with a spike at 110 for the third-born child or more. However, this difference is extremely small compared to the other countries (Sexual Health Team, 2015). In conclusion, the literature supports the assumption of no sex-selective abortions in Colombia.

I reproduce some of the statistics above in the data set and add more information on the children of women who have ever terminated a pregnancy. In the sample consisting of all women who have ever been in a relationship/union/marriage, never had a multiple birth, have non-missing outcomes and at least one living child, the gender ratio is approximately constant throughout the entire birth order. The gender ratios of the second, third, fourth and fifth child are each not statistically significantly different on any relevant significance level from the gender ratio of the first child. The percentage of women who indicated they ever terminated a pregnancy is not statistically significantly different for women whose first-born was a boy versus a girl and for women whose first-born was a girl and whose second-born was a boy versus a girl on any relevant significance level.

Summary statistics

Table 3 and Table 4 show the difference in outcomes and in the extended list of covariates for treatment and control group. Compared to women with a first-born son, women with a first-born daughter are on average two percentage points more likely to have ever experienced domestic violence, which is mostly driven by the difference in less severe domestic violence. The means between treatment and control group are statistically significantly different from each other for domestic violence experienced at least one year prior to the interview. This could be explained by the fact that most of the sons for the women in the sample were born before the last year. Untreated women are three percentage points more likely to have experienced domestic violence, four percentage points more likely to have experienced less severe violence and two percentage points more likely to have been raped by their husband/partner before the last year prior to the interview.

Furthermore, untreated women are more likely to have more children, to live in larger and poorer household in more rural areas and to be older and married. They have a higher degree of autonomy in their relationship, went through less years of schooling and their partners are older. These differences can be partially caused by the gender of their first child, but the way the sample is constructed cannot be neglected as a factor. Based on the son-stopping rule, I expect families whose first-born child was a girl to have more children which will also affect other characteristics of the household. However, differences such as the untreated women being older and more likely to be married point towards the oversampling of young women in the treatment group who are most likely to only have one child, a son, and are therefore included in the sample. These women are more likely to live in urban areas, to have received a higher education, to live in smaller households, to have less children and to not be married. The fact that the treatment and control group are apparently different will be further discussed in Subsection 4.4.

Test of exogeneity

Table 5 shows the results of two regression of the main covariates on the treatment variable to test the exogeneity of the treatment. There is a set of three variable that seems to have a statistically significant impact on the probability of having a first-born son, which does not fully support the assumption that the treatment is exogenous conditional on the set of main variables. The results suggest that older women are less likely to have a first-born son, as well as women who were comparably young when they got their first child and/or who received less education, based on the results with year and department fixed effects.

	Untreated		Treated		p-value
	Mean	N	Mean	N	
Has experienced at least one of the three outcomes					
ever	0.31	15651	0.29	31271	0.00
before last year	0.25	15651	0.22	31271	0.00
during last year	0.19	15651	0.18	31271	1.00
Outcome: Less severe violence					
ever	0.30	15651	0.28	31271	0.00
before last year	0.23	15651	0.19	31271	0.00
during last year	0.17	15651	0.16	31271	1.00
Outcome: Severe violence					
ever	0.11	15651	0.10	31271	0.03
before last year	0.09	15651	0.08	31271	0.00
during last year	0.06	15651	0.06	31271	0.43
Outcome: Sexual violence					
ever	0.07	15651	0.07	31271	0.13
before last year	0.05	15651	0.03	31271	0.00
during last year	0.05	15651	0.04	31271	1.00

Table 3: Difference in outcomes for treatment and control group

The mean and the standard deviation are weighted with the domestic violence sample weight provided by DHS and p-values are adjusted for multiple hypothesis testing using the bonferroni method. The sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. Less severe violence indicates whether the woman has ever experienced at least one of the following items: pushed, shook or had something thrown; slapped; punched with fist or hit by something harmful. Severe violence indicates whether the woman has ever experienced at least one of the following items: kicked or dragged; strangled or burnt; threatened or attacked with weapon. Sexual violence indicates whether the woman was ever physically forced into unwanted sex. All outcomes only refer to actions committed by the husband or partner.

	Untreated		Treated		p-value
	Mean	N	Mean	N	
Characteristics of the household					
number of children 0-5 years	0.73	15651	0.69	31271	0.00
number of children 6-14 years	1.27	15651	0.94	31271	0.00
number of women 13-49 years	1.73	15651	1.40	31271	0.00
number of members	5.25	15651	4.83	31271	0.00
household head is female	0.17	15651	0.19	31271	0.00
lives in urban environment	0.70	15651	0.75	31271	0.00
wealth quintile of household	2.80	15651	2.98	31271	0.00
Characteristics of the female respondent					
experienced domestic violence as child	0.36	15651	0.35	31271	0.58
age difference to partner	4.15	15651	4.34	31271	0.31
age	36.92	15651	34.24	31271	0.00
age at first child	19.67	15651	20.67	31271	0.00
was minor at first sexual intercourse	0.73	15651	0.71	31271	0.00
is divorced	0.00	15651	0.00	31271	0.04
lives with a partner	0.57	15651	0.58	31271	0.87
is married	0.38	15651	0.36	31271	0.00
is separated	0.04	15651	0.05	31271	0.00
is widowed	0.01	15651	0.00	31271	1.00
had first child before getting married	0.08	13016	0.06	26235	0.00
index on autonomy in relationship	2.56	15651	2.49	31271	0.00
is literate	0.69	7318	0.72	10875	0.00
number of dead children	0.11	15651	0.08	31271	0.00
number of living children	3.28	15651	2.44	31271	0.00
number of unions	1.28	15651	1.22	31271	0.00
works in agriculture	0.05	15651	0.04	31271	0.00
works in clerus	0.05	15651	0.07	31271	0.00
works in sales	0.23	15651	0.24	31271	0.09
works in services	0.41	15651	0.37	31271	0.00
works in skilled & manual job	0.04	15651	0.05	31271	1.00
works in technical & managerial job	0.06	15651	0.08	31271	0.00
works in unskilled & manual job	0.02	15651	0.02	31271	1.00
age of partner	41.07	15651	38.57	31271	0.00
is pregnant	0.02	15651	0.03	31271	0.00
is the household head	0.13	15651	0.12	31271	0.00
is wife of household head	0.78	15651	0.73	31271	0.00
ever terminated a pregnancy	0.28	15651	0.26	31271	0.00
total years of education	7.45	15651	8.66	31271	0.00
is currently working	0.54	15651	0.54	31271	1.00

Table 4: Difference in main variables for treatment and control group

The mean and the standard deviation are weighted with the domestic violence sample weight provided by DHS and p-values are adjusted for multiple hypothesis testing using the bonferroni method. The sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. The autonomy index is self-imputed based on five variables that indicate how responsibilities are shared in the relationship and ranges from 0 to 5, whereby 5 indicates that the respondent has full autonomy in all five categories. The variable "experienced domestic violence as child" refers to actions committed by the father against the mother during childhood. "Minor" refers to all women younger than 18 years.

Outcome: Treatment	**Linear probability model**			
Variable	est	p-value	est	p-value
lives in urban environment	0.05	0.00	0.01	1.00
wealth quintile of household	0.12	0.00	0.01	1.00
experienced domestic violence as child	0.06	0.00	-0.01	1.00
age	-0.13	0.00	-0.09	0.00
age at first child	0.07	0.00	0.07	0.00
works in agriculture	0.41	0.00	-0.01	1.00
works in sales	0.25	0.00	0.00	1.00
works in services	0.27	0.00	-0.00	1.00
works in skilled & manual job	0.25	0.00	0.02	1.00
works in technical & managerial job	0.25	0.00	-0.02	1.00
works in unskilled & manual job	0.28	0.00	-0.00	1.00
total years of education	-0.06	0.00	0.02	0.00
age difference to partner	0.01	0.06	0.01	0.04
was minor at first sexual intercourse	-0.01	1.00	0.00	1.00
Fixed effects included	no		yes	

Table 5: Exogeneity of treatment

The sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. The continuous variables are scaled and fixed effects are on a year and department level. Standard errors are robust and clustered as adviced by DHS and p-values are adjusted for multiple hypothesis testing based on the bonferroni method. The observations are weighted with the domestic violence sample weight provided by DHS. The variable "experienced domestic violence as child" refers to actions committed by the father against the mother during childhood. "Minor" refers to all women younger than 18 years.

4.4 Propensity Score

This section shows the extent of the sample selection based on the general assumption that the probability to have a first-born son versus a first-born daughter is exogenous for all women, regardless of their characteristics. This expectation can be tested using the propensity score (Rosenbaum and Rubin; 2006). The propensity score $e(x)$ is the estimated probability that a woman has a first-born son given her observable characteristics:

$$e(x) = E[W_i|X_i = x] \tag{1}$$

If the treatment W is fully randomized with respect to x and percentage p of the sample received the treatment, $e(x) = p$ for all women.

I use a tuned regression forest and predict the propensity scores based on the main covariates.[11] Figure 3 shows a large difference in the propensity scores between women included in the main sample and women who are not. The yellow line demonstrates the estimated propensity score of women who match the sample requirements and have at least one living child. As assumed, the probability to have a first-born son is approximately constant for all women and around 0.5. The estimated propensity score ranges from approximately 0.50 to 0.53 with a mean of 0.511. However, for the women included in the main sample who have at least one living son, the estimated propensity score ranges from approximately 0.6 to 0.8 with a mean of 0.68. This implies that for some women in the sample, the probability to have a first-born son based on the main covariates is around 80 percent, which is unrealistic and has to be attributed to the sample composition.

Table 21 and Table 22 in the Appendix show the difference between all women with at least one living child (N = 78,187) and all women with at least one living son (N = 46,922) in the outcomes and the extended list of covariates. Women with at least one living son are less likely to experience any type of domestic violence during almost any time period. This can be due to the fact that women who are excluded in the main sample have only daughters. This implies the father's preferences for sons, if they exist, were never fulfilled and consequently, his wife/partner experienced an increase in domestic abuse.

Table 6 shows the difference in covariates for women who have a propensity score below or above the 90th percentile of all propensity scores in the main sample with women who have at least one living son. The difference in propensity scores seems to be driven by a combination of variables. In the sample, women who are extremely likely to have a first-born son are likely to live in a less wealthy household in an urban area, to have received more years of schooling, to be younger generally and younger at the birth of their first child. This is probably due to the sample combination, since counterparts with a first-born girl are missing in the main sample for those women who only have one child, a son. Therefore, their characteristics are overly predictive for having a first-born son compared to women with a higher number of children who are more likely to have a counterpart with similar characteristics, but a first-born daughter in the sample. Given the results in Table 6, I oversample women in the treatment group who are young, highly educated, live in cities and are very likely to have only one child.

Figure 4 shows the distribution of the propensity scores of the control and treatment group within the restricted sample of women who have at least one living son. For the control group (N = 15,651), the propensity score ranges

[11] The regression forest algorithm is explained in Appendix Section Additional Explanations to the Causal Forest and Regression Forest Algorithm.

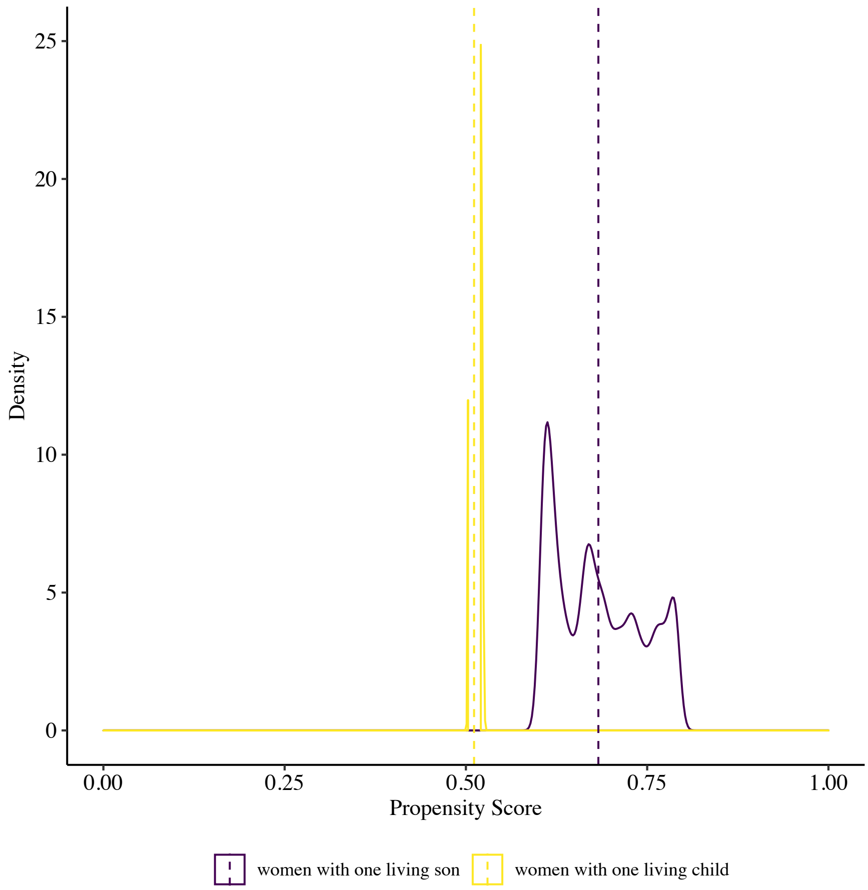

Figure 3: Distribution of propensity score for different samples

The sample is restricted to women who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. The continuous variables are scaled. The propensity score is estimated with a regression forest using tuned parameters (min.node.size is fixed to 1000) and includes fixed effects on a year and department level, clusters as adviced by DHS and the domestic violence sample weight provided by DHS. The blue line shows the propensity score for women who have at least one living son (N = 46,922) and the yellow line shows the propensity score for women who have at least one living child (N = 78,187). The dashed lines indicate the mean of the respective group.

Propensity Score	Below the 90th percentile		Above the 90th percentile		p-value
	Mean	N	Mean	N	
experienced domestic violence as child	0.35	42229	0.36	4693	1.00
lives in urban environment	0.73	42229	0.79	4693	0.00
wealth quintile of household	2.93	42229	2.80	4693	0.00
age difference to partner	-0.04	42229	0.11	4693	0.00
age	0.20	42229	-1.52	4693	0.00
age at first child	0.15	42229	-0.35	4693	0.00
was minor at first sexual intercourse	-0.11	42229	0.38	4693	0.00
works in agriculture	0.04	42229	0.02	4693	0.00
works in clerus	0.06	42229	0.09	4693	0.00
works in sales	0.23	42229	0.28	4693	0.00
works in services	0.39	42229	0.33	4693	0.00
works in skilled & manual job	0.05	42229	0.04	4693	1.00
works in technical & managerial job	0.08	42229	0.06	4693	0.00
works in unskilled & manual job	0.02	42229	0.02	4693	1.00
total years of education	0.04	42229	0.67	4693	0.00
treatment	0.66	42229	0.87	4693	0.00

Table 6: Difference in propensity score

The mean and the standard deviation are weighted with the domestic violence sample weight provided by DHS, the continuous variables are scaled and p-values are adjusted for multiple hypothesis testing using the bonferroni method. The sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. The propensity score is estimated with a regression forest using tuned parameters (min.node.size is fixed to 1000), the domestic violence sample weight provided by DHS, fixed effects on a year and department level and clusters as adviced by DHS. The difference in means is excluded for the variables 'year' and 'department' since they are included as factors in the analysis. The threshold is the 90th percentile of the propensity score, 0.775. The treatment variable is excluded for the estimation of the propensity score. The variable "experienced domestic violence as child" refers to actions committed by the father against the mother during childhood. "Minor" refers to all women younger than 18 years.

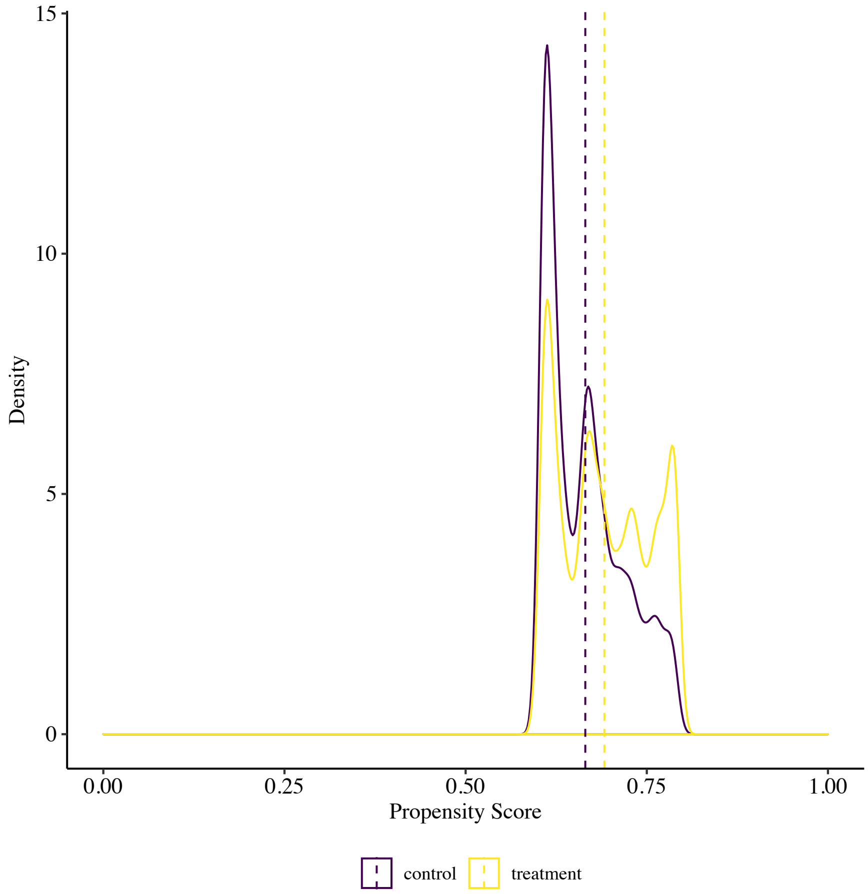

Figure 4: Distribution of propensity score for treatment and control group

The sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. The continuous variables are scaled. The propensity score is estimated with a regression forest using tuned parameters (min.node.size is fixed to 1000) and includes fixed effects on a year and department level, clusters as adviced by DHS and the domestic violence sample weight provided by DHS. The blue line shows women whose first-born child was a girl (N = 15,651) and the yellow line shows women whose first-born child was a son (N = 31,271). The dashed lines indicate the mean of the respective group.

from 0.6 to 0.80 with a mean of around 0.67. In the treatment group (N = 31,271), the propensity score has a very similar range with a mean of 0.69. The mean of the propensity score is statistically significantly different from zero on the 0.01% significance level for treatment and control group. The distribution of the propensity score is shifted towards the right tail of the interval $[0, 1]$ compared to the distribution of the control group.

I can also analyze the results of the regression forest to provide further evidence on the exogeneity of the treatment. Table 23 in the Appendix contains information on the variables used for the splitting decision in the regression forest for the main sample.

The difference in the distribution of the propensity score for control and treatment group is cause to worry when using methods such as the linear probability, logit or probit model for which results are presented in Subsection 6.1. Common ways to deal with the differences in the distribution of the propensity scores between control and treatment group are balancing or matching methods, such as entropy balancing suggested in Hainmueller (2012). The author suggests to recalibrate the sample weights in order to match the distribution of the covariates of control and treatment group.[12] There is no need for any rebalancing when using the causal forest described in Section 5.1 which is doubly-robust and controls for the propensity score (Athey and Wager, 2019).

5 Methodology: Heterogeneous Treatment Effects

Why are heterogeneous treatment effects relevant?

There is a broad range of settings for which the analysis of heterogeneous treatment effects is crucial, such as medical trials. The same medication is very likely to have different effects on certain subgroups, whereby the subgroups can be defined over observable characteristics, such as gender or age. When deciding which drug to give a patient, it is important to take into account that this patient is very likely different from the average patient in these observable characteristics and therefore, the treatment effect the patient can expect from the drug is potentially different compared to the average treatment effect (Willke et al., 2012).

In Economics, similar examples can among others be found when evaluating public policies, such as tax reforms. The reform can have different effects on different subgroups of the populations and one might be interested in considering the heterogeneous treatment effects when designing the reform. Furthermore, in combination with restrictive budget constraints, identified heterogeneous treatment effects allow to assign the treatment to subpopulations that are expected to benefit the most.

Why is there a need for heterogeneous treatment effect methods?

One might be tempted to think that the identification of heterogeneous treatment effects does not require new methods, since ways to look for these effects already exist in the literature. Popular methods consist of rerunning the regression design on specific subsets of the sample or to include interaction terms between the treatment and covariates. However, several problems come up with these identification strategies which are due to the setup of the data and testing procedures.

Randomized control trials and similar experiments are often run in the following way:

1. A pre-analysis plan has to be submitted in which all the planned steps of analysis are included. If the existence of heterogeneous treatment effects is anticipated, the subgroups based on the values of certain covariates have to be specified for which different treatment effects are expected. However, in reality this is often difficult due to limited prior knowledge of the treatment effect.
2. After the experiment is conducted, the analysis is restricted to the pre-specified covariates, such as gender, for which treatment effect heterogeneity is expected. Unexpected heterogeneity for other covariates cannot be analyzed, even if it is in the data, since it is difficult to justify the selection of specific subgroups with heterogeneous treatment effects *ex-post* due to data mining concerns (Wager and Athey, 2018). There are several concerns that come up:

[12] I have implemented the entrophy balancing for the data set, based on the distribution of the main variables listed in the output tables in Section 6.1. However, in this specific setting, the domestic violence sample weights provided by DHS seem to be sufficient to generate a similar distribution of these covariates and eliminates the need to recalibrate the sample weights. This could among others be due to the low number of variables compared to the number of observations in control and treatment group.

(a) The treatment and control group cannot be compared in the variable that defines the subgroups with heterogeneous treatment effects, since the randomization with respect to the variable was not conducted in a way that allows to determine heterogeneous treatment effects. See Imai and Ratkovic (2013) as example for a paper that treats the identification problem of heterogeneous treatment effects based on subgroups as one of variable selection.
(b) When dividing the sample into subgroups based on the value of the variable, the number of observations is too small to actually identify heterogeneous treatment effects. More generally, lack of power can be a concern to detect heterogeneous treatment effects.
(c) When trying potential combinations of variables to determine subgroups with heterogeneous treatment effects, multiple hypothesis testing and its consequences have to be taken into account (Willke et al., 2012). For example, when using a significance level of 1% and using 200 different combinations of covariate values to split the sample in subsamples, we expect around two treatment effects to be statistically significant from zero even though there are actually no valid treatment effects in the sample. This is due to false positives. See Lee and Shaikh (2014) as example for a paper that treats the identification of heterogeneous treatment effects based on subgroups ex-post as a multiple hypothesis problem.

Sophisticated heterogeneous treatment effects methods allow a data-driven search for heterogeneity in causal effects with valid standard errors. Instead of having to pre-specify certain subgroups for which heterogeneous treatment effects are expected, only a set of covariates has to be given among which the methods attempt to identify different treatment effects. This thesis relies on the causal forest algorithm implemented in the R-package grf.

5.1 Introduction to Causal Forests

Notation

The notation used in the following relies on the potential outcomes framework introduced by Rubin (1974):

X_i a p-dimensional vector of observable pre-treatment characteristics for woman i
$W_i \epsilon \{0,1\}$ a binary variable indicating whether woman i was treated (1) or not (0)
$Y_i^{obs} \epsilon \{0,1\}$ a binary variable indicating whether woman i experienced domestic violence
$Y_i\,(1)$ the outcome Y woman i would have attained if she had been treated
$Y_i\,(0)$ the outcome Y woman i would have attained if she had not been treated

Using the potential outcome notation, Y_i^{obs} can be written as:

$$Y_i^{obs} = W_i Y_i(1) + (1 - W_i) Y_i(0) \tag{2}$$

The average treatment effect (ATE) is the expected value of the difference in outcomes between the treatment and control group if the treatment was fully randomized:

$$\tau = E[Y_i(1) - Y_i(0)] \tag{3}$$

The conditional average treatment effect (CATE) is the difference in outcomes between the treatment and control group, conditional on the set of covariates x. The CATE estimates the ATE in a subgroup of the sample, defined over the values of the covariates and therefore allows to identify heterogeneous treatment effects:

$$\tau(x) = E[Y_i(1) - Y_i(0)|X_i = x] \tag{4}$$

The expected outcome marginalizing over treatment is the expected probability to experience domestic violence given the covariate values:

$$m(x) = E[Y_i|X_i = x] \tag{5}$$

The expected outcome in Equation 5 is not equal to the expected outcome in Equation 6 if we expect the gender of the first child to have an impact on the probability to experience domestic violence:

$$\mu_{(w)}(x) = E[Y_i|X_i = x, W_i = w] \tag{6}$$

Assumptions

The two following main assumptions have to be fulfilled when using the causal forest or any other method aiming to estimate causal effects:

1. Unconfoundness: Once we know all observable characteristics X_i of woman i, then knowing about whether the woman was treated $W_i = 1$ or not $W_i = 0$ provides us with no additional information about her *a priori* probability of experiencing domestic violence due to unfulfilled preferences for sons. This implies that the extent of the husband's/partner's preferences of sons and his resulting threshold of abusing his wife/partner due to unfulfilled preferences of sons does not impact the probability of having a first-born son. The gender of the first child is also randomly assigned within each subpopulation of women, defined over $X_i = x$.

$$Y_i(1), Y_i(0) \perp W_i|X_i \tag{7}$$

   Equation 7 is not fulfilled if women whose partner is more likely to react violently to a first-born girl are more likely to give birth to a son or vice versa. This could for example be caused by non-random sex-selective abortions by women who are more likely to experience domestic violence due to unfulfilled preferences of sons.

2. Overlap: There is no subpopulation among the women indexed by a certain set of covariates $X_i = x$ that is either completely in the treatment or control group. $\forall x \epsilon supp(X)$

$$0 < P(W = 1|X = x) < 1 \tag{8}$$

   Equation 8 states the propensity score for each set of covariates has to be larger than 0 and smaller than 1 and implies that every woman included in the main sample has to have an *a priori* probability of having a first-born son that is larger than 0 and smaller than 1. For example, this is violated if the number of children is included as covariate in the analysis. All women with only one child in the main sample are included in the treatment group due to the setup of the identification strategy and therefore, the propensity score for all women with one child, regardless of the value of the other variables, is 1.

Given the assumptions mentioned above, causal forests can consistently estimate the true treatment effect (Wager and Athey, 2018).

Difference to general Machine Learning procedures

In general, Machine Learning methods result in predictions. The goal of predictions is to estimate $m(x) = E[Y|X = x]$ and to predict Y in a new data set where only x is observed. The goal function which is minimized by the final model is the mean square error (MSE):

$$MSE \equiv \frac{1}{i}\sum_{i}(Y_i - \hat{m}(X_i))^2$$

Supervised methods perform the following steps:

1. The original data set is randomly divided in a training and test data set.
2. The model is trained on the training set.
3. The performance of the model is evaluated using the test set, where we still observe the original Y. Predictions based on x, for which the observation was not used to train the model are referred to as *out-of-bag*.

Machine Learning methods that are focused on making predictions are not suited to estimate causal effects due to bias concerns. When minimizing the MSE, there is a bias-variance trade-off and the best model will always include some bias. The trade-off describes the fact that if the estimators are too sensitive to the original data set and therefore, perform very well within the data set, they perform poorly on a new data set, which is referred to as overfitting. In general, the higher the bias, the worse is the in-sample prediction on the training set. The higher the variance, the better the in-sample prediction on the training set, which leads to overfitting.

When attempting to estimate heterogeneous treatment effects with Machine Learning methods, there are two main differences to the procedure described above:

First, the treatment we are trying to predict is unknown in the training and test set, since we never observe $Y_i(0)$ and $Y_i(1)$ for the same individual. The causal forest relies on the assumptions mentioned above (Wager and Athey, 2018). The unobserved value for Y_i can then be inferred based on observations with similar covariate values.

Second, the goal is to receive unbiased estimates. The causal forest relies on the honesty principle to achieve this, which will be explained in Section 5.2.

5.2 Underlying Principles of Causal Forests

Trees and Forests

In general, trees split the original data set in subgroups based on the values of their observable characteristics. For decision trees, the goal is to define subgroups that are as different as possible in terms of the outcome, so that all observations in a subgroup have similar outcomes. The algorithm starts with the entire training sample and the goal for the first step is to maximize the difference in the outcomes between the two resulting subgroups. Ideally, one would use every value of every covariate in the data set to generate two subgroups and calculate the difference in outcomes to identify the maximum difference and the corresponding value-covariate pair. In practice, the algorithm tries several values of several covariates, which are picked randomly and uses the value-covariate pair that results in the largest difference in outcomes between the subgroups. After each split, the data set is divided into two subgroups and the subsequent splits are based on each of the two subgroups individually. In general, the tree will stop splitting once the maximized difference in outcomes for the next split is below a certain threshold.

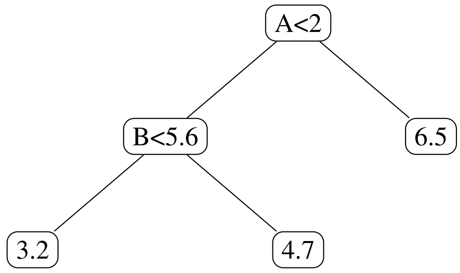

Figure 5: Example for decision tree

Once the tree is built, the test set can be used to predict the outcomes based on the values of the covariates. Each observation is categorized into a leaf following the splitting rules determined before and its predicted outcome is the mean outcome of the observations in the training set that fell in the same leaf.

Figure 5 shows an example of a basic decision tree in a setting with several covariates, among them the continuous variables A and B. There are three final subgroups with the predicted values 3.2, 4.7 and 6.5, which are referred to as terminal nodes. The internal nodes are the ones with the values $A < 2$ and $B < 5.6$. The interpretation of the tree is the following: In the given trainings data set, the value-covariate pair for A and the value 2 results in the largest difference in outcomes in the resulting subgroups. If $A > 2$ for a observation, its predicted outcome is 6.5. If $A < 2$ is true, the next splitting decision is on whether $B < 5.6$ is true for the same observation. If $B < 5.6$ is true, the predicted outcome is 3.2, if it is false the prediction is 4.7.

Decision trees are a useful tool in a setting with a highly non-linear and complex relationship, however, they are not very robust which can lead to a high variance in the final trees, depending on the selection of the training sample and the value-covariate pairs considered for each splitting decision. Forests overcome this weakness, since they are based on building multiple trees. The final predictions are obtained by averaging the predicted values across the trees. Essentially, the estimation process is built on two steps: First, n trees are grown individually. Second, the predictions made for each observation in each of the n trees is averaged over n to get the final prediction for each observation. For classification with a binary outcome, the final prediction is based on how the majority of the trees classified the observation.

As explained above, forests are usually seen as an average of trees. However, in the context of causal forests which will be used in the following, it is more appropriate to interpret them as a weighting approach to estimate the treatment effect. The trees used to build the causal forest are splitting based on the value-covariate pair that maximizes the average treatment effect in the resulting subgroups, which will be explained in detail in Section 5.3. Following the explanation above, a causal forest would simply average out the predicted treatment effect across all trees for each observation. But some leaves can be highly variable across all trees in a causal forest, meaning that the values of covariates x in a leaf can vary strongly over the trees and averaging across these variable leaves is not optimal. When estimating the treatment effect for given values for the covariates $X_i = x$, the causal forest gives more weight to observations in the data that fall more often in the same leaf as the observation with the given values of the covariates across the trees in the forest. These observations are considered close to the one with the given values for the covariates. This is also helpful to determine which covariates are important among the set of all covariates for the closeness of two observations, since the leafs are the results of splitting decisions based on covariates. If two observations are close, they must have similar values for the variables that are used as split criterion and therefore, these variables are determinants of closeness.

Doubly-robustness

Under the assumption of unconfoundness, the average treatment effect can be written as

$$\tau = E[\mu_1(X_i) - \mu_0(X_i)] = E\left[\frac{W_i Y_i}{e(X_i)} - \frac{(1-W_i)Y_i}{1-e(X_i)}\right] \tag{9}$$

and it is easy to see why, if either both $\mu_{(w)}(x)$ or $e(x)$ are estimated perfectly, we receive an unbiased estimator (Wager and Athey, 2018). However, in reality, we never estimate the conditional mean function or the propensity score perfectly due to misspecifications, which can lead to bias in the estimation of the treatment effect. This becomes especially important if there are differences between the covariate distribution in control and treatment group or in the covariate values, which can lead to different propensity score estimates for treatment and control group and similarly to different coefficient estimates for the set of covariates x for $\mu_1(x)$ and $\mu_0(x)$ since they are estimated separately. The difference $\mu_1(x) - \mu_0(x)$ would then not only estimate the treatment effect, but also mirror random differences in the covariate values and distribution between treatment and control group (Powers et al., 2018).

Doubly robust methods combine the models of outcome regression $\mu_{(w)}(x)$ and propensity score $e(x)$ to reduce the sensitivity to misspecifications. Even if only one of the two is well specified, the resulting estimator is much more robust. In the first step of the algorithm, the causal forest estimates $e(x)$ and $m(x)$ to reduce potential bias. This makes the causal forest also suited when dealing with observational data, where the treatment is not assigned randomly.

Honesty

A method is called honest if it uses one part of the data set to estimate the parameters of the model and the remaining part of the data set to produce estimates given the estimated parameters. Below are the exemplary steps for honest causal trees (Athey and Imbens, 2016; Wager and Athey, 2018):

1. The data set is split in two parts. This data set is usually not the original data set, but the training set sampled from the original one.
2. The first part is used to construct the tree. The response Y_i is used to decide where to split, given W_i and X_i.
3. The second part is used to estimate the treatment effects within the leaves of the tree based on the response Y_i. The observations are automatically divided in the leaves based on the splitting decisions resulting from the step before.

The splitting decisions made in Step 2 can be seen as exogenously given in Step 3 for each tree. The estimates resulting from an honest causal forest estimates are unbiased and have valid confidence intervals. This comes at the price of precision loss due to an artificially reduced sample size when estimating the model parameters and the treatment effects. However, the benefit is a reduction or elimination of bias in the resulting treatment effect estimates (Athey and Imbens, 2016).

5.3 Estimation Procedure of Causal Forests

The estimation of heterogeneous treatment effects based on the causal forest algorithm consists mainly of the steps described below in a simplified way (Wager and Athey, 2018; Athey and Wager, 2019; Athey et al., 2019).[13] The

[13] See the website: https://grf-labs.github.io/grf/REFERENCE.html for an introduction to the grf package.

causal forest is used to calculate the weights $\alpha_i(x)$, which are included in the estimation of the treatment effect for each observation with $X_i = x$.

1. The response functions $\hat{e}(x)$ and $\hat{m}(x)$, based on Equations 1 and 5, are estimated separately using the regression forest algorithm
2. Each causal tree in the causal forest is estimated honestly and results in an estimate $\hat{\tau}_b(x)$ for an observation with $X_i = x$ (Athey and Imbens, 2016). Overall, B causal trees are generated. Each split in each tree is determined in the following way:
   (a) The training sample is split in two parts, one for constructing the tree and the second one to estimate the treatment effects within the leaves. This step is repeated for each tree so that the trees are trained on different subsamples of the training sample.
   (b) The splitting rules are determined using only the first part of the training sample in the following way: For each potential "left-right" split (L, R), based on all value of a subsample of the covariates, $\hat{\tau}_L$ and $\hat{\tau}_R$ are computed whereby a constant treatment effect is assumed across all observations in the resulting subgroup:

$$\hat{\tau}_L \leftarrow lm\left(\left(Y_i - \hat{m}^{(-i)}(X_i\right) \sim \left(W_i - \hat{e}^{(-i)}(X_i)\right) : X_i \epsilon L\right)$$

$$\hat{\tau}_R \leftarrow lm\left(\left(Y_i - \hat{m}^{(-i)}(X_i\right) \sim \left(W_i - \hat{e}^{(-i)}(X_i)\right) : X_i \epsilon R\right)$$

   The value-covariate pair which maximizes the weighted difference $n_L n_R(\hat{\tau}_L$ -$\hat{\tau}_R)^2$ is chosen for the splitting decision, whereby n_L is the number of observations that fall in the left node and n_R is the number of observations that fall in the right node. There are no additional splits when no nodes can be split further, which can be set by, among others, the minimum number of observations per leaf, a threshold for the weighted difference in the resulting treatment effects or a difference in sizes between the resulting nodes in terms of control and treatment group.
   (c) The second part of the training sample is used to estimate the treatment effect. The estimation of the treatment effect in leaf L for any $x \epsilon L$ is the difference in outcomes for treatment and control group in each leaf $\bar{y}_1 - \bar{y}_0$:

$$\hat{\tau}(x) = \frac{1}{|\{i : W_i = 1, X_i \epsilon L\}|} \sum_{\{i:W_i=1,X_i\epsilon L\}}^{Y_i} - \frac{1}{|\{i : W_i = 0, X_i \epsilon L\}|} \sum_{\{i:W_i=0,X_i\epsilon L\}}^{Y_i} \tag{10}$$

   The underlying assumptions are that in each leaf, the treatment effect is the same across all observations in the leaf and the leaves are small enough that the (Y_i, W_i) pairs of each observation in a leaf behave as they have come from a randomized experiment. So W_i is randomly distributed across the observations in the leaf based on the covariates and Y_i is not predictive of the treatment status.
3. The learned adaptive weights $\alpha_i(x)$ are calculated based on Equation 11, whereby B is the total number of trees in the forest and $L_b(x)$ is the leaf where x falls into in tree b. The weights indicate how often an observation i in the training sample falls in the same leaf as the observation with the given values of the covariates across the trees in the forest. The more often the observations are in the same leaf, the closer they are to each other and the higher is the weight of observation i when estimating the treatment effect of the

observation with the given values of the covariates based on the outcomes in the training sample.

$$\alpha_i(x) = \frac{1}{B}\sum_{b=1}^{B} \frac{(\{X_i \epsilon L_b(x), i\epsilon B\})}{|\{i : X_i \epsilon L_b(x), i\epsilon B\}|} \tag{11}$$

4. If we see the causal forest as an average of B causal trees, the estimated treatment effect is the average of the sum of all estimated treatment effects across each causal tree as described in Equation 12:

$$\hat{\tau}(x) = B^{-1}\sum_{b=1}^{B} \hat{\tau}_b(x) \tag{12}$$

More precisely, the treatment effect for each observation is estimated as described in Equation 13, using the weights $\alpha_i(x)$ obtained through the causal forest:[14]

$$\hat{\tau}(x) = \frac{\sum_{i=1}^{n} \alpha_i(x)(Y_i - \hat{m}^{(-i)}(X_i))(W_i - \hat{e}^{(-i)}(X_i))}{\sum_{i=1}^{n} \alpha_i(x)(W_i - \hat{e}^{(-i)}(X_i))^2} \tag{13}$$

$\hat{m}^{(-i)}$ and $\hat{e}^{(-i)}$ are out-of-bag estimates, meaning that Y_i was not used to estimate $\hat{m}^{(-i)}(X_i)$ and $\hat{e}^{(-i)}(X_i)$. The estimated treatment effect for an observation with the covariate values x is based on the mean of the outcomes in the training sample, weighted with the weights calculated in the causal forest and the residual treatment and outcome.

Additional features of the causal forest algorithm, such as tuning and the estimation of the average treatment effect and the algorithm of regression forest used in Step 1 are explained in Appendix Section Additional Explanations to the Causal Forest and Regression Forest Algorithm.

6 Results

6.1 Average Treatment Effect: LPM, Probit and Logit

Table 7, Table 8 and Table 9 show the results for each of the domestic violence outcomes that the women have ever experienced. Table 24, Table 25 and Table 26 in the Appendix show the same results for domestic violence that occurred before the last year. Table 27, Table 28 and Table 29 in the Appendix show the same results for domestic violence that occurred during the last year.

Across most outcomes and time periods we observe that growing up in an urban, poorer household with domestic violence as a child, being younger and being younger when giving birth to the first child and working in services or sales increases the probability for women to experience domestic violence by their husband/partner. Previous literature has demonstrated that occupational choice is heavily influenced by personality (see among others Ackerman and Beier, 2003; Tracey and Hopkins, 2001). Therefore, the results with respect to the profession should be interpreted as rough indicator for personality traits, rather than the actual work in the specific sector.

[14]Recall that in OLS, the estimator of the slope $\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sum_{i=1}^{n}(X_i-\bar{X})^2}$. Compared to this, the causal forest estimator can been seen as a weighted approach, with W as the variable of interest, controlled for the expected marginalized outcome and the propensity score

Less severe violence

Additionally to the results above, women are more likely to have experienced less severe violence when they were minor at their first sexual intercourse. The coefficient of the treatment dummy has the sign expected from preferences for sons and is statistically significantly different from zero in the probit model across all time periods on the 1% significance level and additionally in the logit model for the ever experienced domestic violence and violence experienced before the last year prior to the interview on the 5% significance level. Women whose first-born child is a son are less likely to have experienced less severe domestic violence. Women are more likely to have experienced less severe domestic violence during the last year if they younger, whereas older women are more likely to have experienced this type of violence before the last year. A smaller difference in age between the women and her husband/partner and working in a skilled and manual job increase the probability of having experienced domestic violence during the last year and having ever experienced less severe violence. Women working in agriculture also face an increased probability of having experienced domestic violence before the last year and having ever experienced less severe violence.

Severe violence

Additionally to the results above, younger women, those who received less education and those with a smaller age difference to their husband/partner are more likely to have experienced severe domestic violence during the year before the interview. Older women and those who were minor at their first sexual intercourse are more likely to have experienced severe domestic violence before the last year prior to the interview or ever. There is no striking evidence for an effect of the treatment variable on the probability of having experienced severe domestic violence.

Sexual violence

Additionally to the results above, older women are more likely to have been raped by their husband or partner. There is no striking evidence for an effect of the treatment variable on the probability of having experienced sexual domestic violence.

	Linear probability model			**Logit**			**Probit**		
Variables	est	s.e	p-value	est	s.e	p-value	est	s.e	p-value
lives in urban environment	0.08	0.01	0.00	0.32	0.03	0.00	0.32	0.02	0.00
wealth quintile of household	-0.03	0.00	0.00	-0.11	0.01	0.00	-0.11	0.01	0.00
experienced domestic violence as child	0.12	0.01	0.00	0.57	0.02	0.00	0.57	0.01	0.00
age	-0.00	0.00	1.00	-0.02	0.01	1.00	-0.02	0.01	0.07
age at first child	-0.03	0.01	0.00	-0.19	0.02	0.00	-0.19	0.01	0.00
works in agriculture	0.05	0.01	0.01	0.31	0.05	0.00	0.31	0.03	0.00
works in sales	0.06	0.01	0.00	0.31	0.04	0.00	0.31	0.02	0.00
works in services	0.05	0.01	0.00	0.35	0.03	0.00	0.35	0.02	0.00
works in skilled & manual job	0.07	0.02	0.00	0.47	0.06	0.00	0.47	0.04	0.00
works in technical & managerial job	0.00	0.01	1.00	0.10	0.06	1.00	0.10	0.03	0.07
works in unskilled & manual job	0.06	0.02	0.20	0.32	0.08	0.00	0.32	0.05	0.00
total years of education	-0.00	0.00	1.00	-0.00	0.02	1.00	-0.00	0.01	1.00
age difference to partner	-0.02	0.00	0.00	-0.08	0.01	0.00	-0.08	0.01	0.00
was minor at first sexual intercourse	0.02	0.00	0.00	0.10	0.01	0.00	0.10	0.01	0.00
treatment	-0.01	0.01	0.95	-0.09	0.02	0.01	-0.09	0.01	0.00

Table 7: Regression results: Less severe violence

The sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. The continuous variables are scaled and fixed effects are on a year and department level. Standard errors are robust and clustered as adviced by DHS and p-values are adjusted for multiple hypothesis testing based on the bonferroni method. The observations are weighted with the domestic violence sample weight provided by DHS. Less severe violence indicates whether the woman has ever experienced at least one of the following items: pushed, shook or had something thrown; slapped; punched with fist or hit by something harmful. The outcome only refers to actions committed by the husband or partner. The variable "experienced domestic violence as child" refers to actions committed by the father against the mother during childhood. "Minor" refers to all women younger than 18 years.

	Linear probability model			**Logit**			**Probit**		
Variables	est	s.e	p-value	est	s.e	p-value	est	s.e	p-value
lives in urban environment	0.04	0.01	0.00	0.34	0.05	0.00	0.34	0.02	0.00
wealth quintile of household	-0.02	0.00	0.00	-0.16	0.02	0.00	-0.16	0.01	0.00
experienced domestic violence as child	0.05	0.00	0.00	0.56	0.03	0.00	0.56	0.02	0.00
age	0.01	0.00	0.00	0.14	0.02	0.00	0.14	0.01	0.00
age at first child	-0.02	0.01	0.08	-0.26	0.02	0.00	-0.26	0.01	0.00
works in agriculture	0.02	0.01	1.00	0.32	0.07	0.00	0.32	0.04	0.00
works in sales	0.04	0.01	0.00	0.38	0.06	0.00	0.38	0.03	0.00
works in services	0.03	0.00	0.00	0.47	0.05	0.00	0.47	0.02	0.00
works in skilled & manual job	0.04	0.01	0.01	0.62	0.09	0.00	0.62	0.05	0.00
works in technical & managerial job	0.00	0.01	1.00	0.15	0.09	1.00	0.15	0.05	0.05
works in unskilled & manual job	0.04	0.01	0.04	0.47	0.11	0.00	0.47	0.06	0.00
total years of education	-0.01	0.00	0.08	-0.10	0.02	0.00	-0.10	0.01	0.00
age difference to partner	-0.00	0.00	1.00	-0.04	0.02	0.40	-0.04	0.01	0.00
was minor at first sexual intercourse	0.01	0.00	0.04	0.13	0.02	0.00	0.13	0.01	0.00
treatment	0.00	0.00	1.00	-0.06	0.03	1.00	-0.06	0.02	0.03

Table 8: Regression results: Severe violence

The sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. The continuous variables are scaled and fixed effects are on a year and department level. Standard errors are robust and clustered as adviced by DHS and p-values are adjusted for multiple hypothesis testing based on the bonferroni method. The observations are weighted with the domestic violence sample weight provided by DHS. Severe violence indicates whether the woman has ever experienced at least one of the following items: kicked or dragged; strangled or burnt; threatened or attacked with weapon. The outcome only refers to actions committed by the husband or partner. The variable "experienced domestic violence as child" refers to actions committed by the father against the mother during childhood. "Minor" refers to all women younger than 18 years.

	Linear probability model			**Logit**			**Probit**		
Variables	est	s.e	p-value	est	s.e	p-value	est	s.e	p-value
lives in urban environment	0.02	0.01	0.00	0.27	0.05	0.00	0.27	0.03	0.00
wealth quintile of household	-0.01	0.00	0.00	-0.15	0.02	0.00	-0.15	0.01	0.00
experienced domestic violence as child	0.04	0.00	0.00	0.57	0.04	0.00	0.57	0.02	0.00
age	0.02	0.00	0.00	0.26	0.02	0.00	0.26	0.01	0.00
age at first child	-0.01	0.00	0.00	-0.24	0.03	0.00	-0.24	0.01	0.00
works in agriculture	0.03	0.01	0.01	0.37	0.09	0.00	0.37	0.04	0.00
works in sales	0.02	0.00	0.00	0.28	0.06	0.00	0.28	0.03	0.00
works in services	0.02	0.00	0.00	0.36	0.06	0.00	0.36	0.03	0.00
works in skilled & manual job	0.04	0.01	0.05	0.48	0.10	0.00	0.48	0.05	0.00
works in technical & managerial job	0.00	0.01	1.00	0.12	0.11	1.00	0.12	0.05	0.76
works in unskilled & manual job	0.01	0.01	1.00	0.28	0.13	1.00	0.28	0.06	0.00
total years of education	-0.01	0.00	0.31	-0.03	0.03	1.00	-0.03	0.01	1.00
age difference to partner	0.00	0.00	1.00	0.01	0.02	1.00	0.01	0.01	1.00
was minor at first sexual intercourse	0.01	0.00	0.05	0.09	0.03	0.02	0.09	0.01	0.00
treatment	0.00	0.00	1.00	-0.00	0.04	1.00	-0.00	0.02	1.00

Table 9: Regression results: Sexual violence

The sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. The continuous variables are scaled and fixed effects are on a year and department level. Standard errors are robust and clustered as adviced by DHS and p-values are adjusted for multiple hypothesis testing based on the bonferroni method. The observations are weighted with the domestic violence sample weight provided by DHS. Sexual violence indicates whether the woman was ever physically forced into unwanted sex. The outcome only refers to actions committed by the husband or partner. The variable "experienced domestic violence as child" refers to actions committed by the father against the mother during childhood. "Minor" refers to all women younger than 18 years.

6.2 Heterogeneous Treatment Effects: Causal Forest

General interpretation of causal forest output

There are several things to keep in mind when interpreting the output of causal forests. First, there is not one systematic measure that allows to evaluate the output of causal forests. Instead several pieces of analysis have to be combined to paint a clearer picture. Second, when describing the subgroups for which the treatment effect differs in size, there are several ways or several sets of variables to do so. Especially when two variables are highly correlated, the one that ends up being selected in each tree for the splitting rule does probably not affect the definition of the subgroups. Based on this, the chosen covariates are just one way to describe the subgroups and are characteristics of the subgroups.

Results

In the following, I will only present the results of one outcome variable in-depth, which is whether the woman ever experienced less severe violence. As mentioned above, there is not a single measure that can be interpreted as pure evidence for heterogeneity in treatment effects, instead the analysis is more complex. Based on the entire output generated for each of the outcomes, I do not find valid heterogeneity in treatment effects for severe or sexual violence. This does not imply that there is no actual heterogeneity as explained below.[15] The evidence for heterogeneity in treatment effects for less severe violence is most compelling for whether the woman experienced that type of violence ever or during the last year.[16] I present a selection of the plots and tables for the last-mentioned outcome in Appendix Section Additional Results for Heterogeneous Treatment Effects.

Figure 6 shows the distribution of the predicted out-of-bag CATE estimates of the causal forest. It is generally difficult to use the range of the predicted values as sole indication for heterogeneous treatment effects. The range of values can be, among others, impacted by potential overfitting or by a lack of power to detect heterogeneity. Based on the range of values in Figure 6, the effect of having a first-born son seems to slightly decrease the probability of having experienced less severe domestic violence for most women in the sample. For some women, the probability seems to slightly increase when having a first-born son.

In the following, the sample is split in four quartiles, based on the value of the predicted treatment effect to simplify the analysis of the causal forest output. I will look at how these quartiles are different in the size of the predicted treatment effect and in the values of the covariates to get an impression of which variables are related to heterogeneity.

Table 10 compares the average treatment effect across the quartiles, estimated with two different methods. In each quartile, the sample ATE is the difference of the average outcome between treatment and control group. The augmented inverse-propensity weighted (AIPW) ATE is estimated doubly-robust using Equation 15 in the Appendix and takes into account the conditional mean outcome and the estimated propensity score. Therefore, the AIPW ATE results most likely in less biased estimates of the treatment effect. The sample ATEs across the quartiles are all negative and not statistically significantly different from each other based on the p-value. The

[15] Due to space concerns, I decided not to include the entire output for each outcome in the main thesis or appendix. Please feel free to reach out to me directly if you are interested.

[16] This is not contrary to the hypothesis that I expect the treatment effect to be strongest in size for the time periods "ever" and "before last year" (based on the assumption that most sons were born before the last year in the sample) since the heterogeneity in treatment effect does per se not give information on the size of the average treatment effect. Rather, the average treatment effect for the outcome "before last year" can be larger in size than the highest treatment effect for a defined subgroup for "during last year". The defined subgroup could consist of women who gave birth to their first son last year.

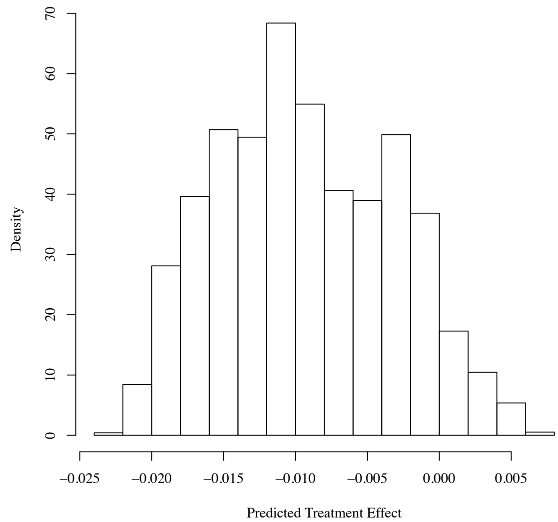

Figure 6: Main results: Distribution of out-of-bag CATE

The x-axis shows the distribution of predicted treatment effects for the training sample (N = 37,537) and the y-axis the probability densities. The causal forest is estimated using 0.8 of the sample as training sample with input parameters tuned once on the training sample and the domestic violence sample weight provided by DHS. The number of trees equals the number of women in the training sample. The sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. Less severe violence indicates whether the woman has ever experienced at least one of the following items: pushed, shook or had something thrown; slapped; punched with fist or hit by something harmful. The outcome only refers to actions ever committed by the husband or partner.

	Sample ATE	AIPW ATE
Quartile 1	-0.0273	-0.0398
	(0.0098)	(0.0113)
Quartile 2	-0.0264	-0.0086
	(0.0098)	(0.011)
Quartile 3	-0.0356	-0.0155
	(0.0098)	(0.0098)
Quartile 4	-0.0179	0.0249
	(0.01)	(0.0104)
p-value	0.6569	3e-04

Table 10: Main results: ATE estimates within quartiles of treatment effect

The quartiles on the y-axis are the quartiles of the treatment effect, the ATE per subgroup is defined by the out-of-bag CATE. The sample ATE is the difference of the average outcome between control and treatment group in the treatment effect quartile. The AIPW ATE is the augmented inverse-propensity weighted average treatment effect in the treatment effect quartile. The p-values are testing the null hypothesis: ATE is constant across the quartiles. The sample ATE uses an F-Test and the AIPW ATE uses a Wald test. The causal forest is estimated using 0.8 of the sample as training sample with input parameters tuned once on the training sample and the domestic violence sample weight provided by DHS. The number of trees equals the number of women in the training sample (N = 37,537). The sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. Less severe violence indicates whether the woman has ever experienced at least one of the following items: pushed, shook or had something thrown; slapped; punched with fist or hit by something harmful. The outcome only refers to actions ever committed by the husband or partner.

sample treatment effects are statistically significantly different from zero on the 1% significance level in the first three quartiles. The estimated AIPW ATE is statistically significant from zero on the 1% significance level in quartile one and four, and is statistically significantly different across the quartiles of the predicted treatment effect on the 0.1% significance level. In the first quartile, women with a first-born son are four percentage points less likely to have experienced less severe domestic violence compared to women with a first-born girl. However, women in the fourth quartile are 2.5 percentage points more likely to have experienced less severe violence when having a first-born son.

Figure 7 shows the estimated average treatment effect per quartile of the predicted treatment effect for the two different estimation methods. It can be inferred easily that the methods result in similar estimates of the average treatment effect in the first three quartiles, but the estimates for the fourth quartile are statistically significantly different on the 5% significance level. The treatment effect of the fourth quartile seems to drive the difference in p-values in Table 10. Apparently, the expected outcome marginalized over the treatment and the estimated propensity scores impact the estimation of the average treatment effect, especially in the fourth quartile. This questions the validity of the results presented in Section 6.1, since, contrary to the causal forest, methods such as the LPM, logit and probit do not take into account the differences in propensity scores between treatment and control group.

Similar to the results in Section 6.1, the AIPW ATEs in the entire training sample, on the treated and the untreated in the training sample based on the causal forest estimation are very similar and not statistically significantly different from zero on any relevant significance level. Therefore, Figure 7 also demonstrates the advantages when splitting the sample in quartiles based on predicted treatment effects which allows to detect statistically significant treatment effects in subgroups that were determined by the algorithm.

Table 11 shows the difference in the AIPW ATE estimates between the four quartiles of the predicted treatment effect. The difference in the AIPW ATE between the first and the fourth quartile is 6.5 percentage points and

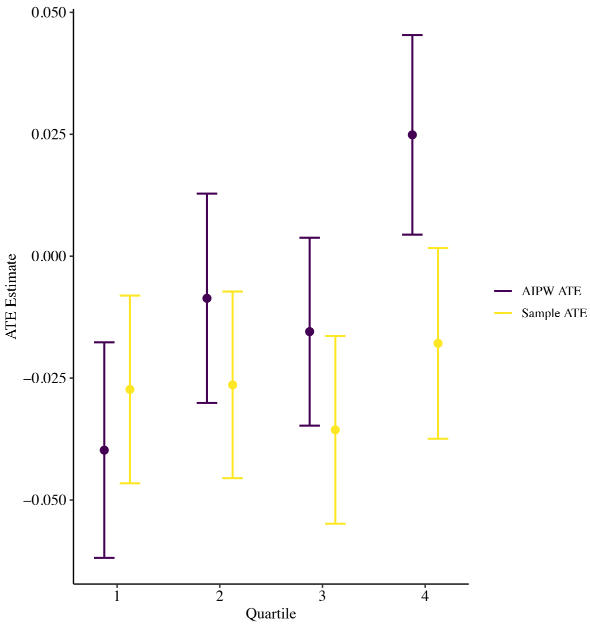

Figure 7: Main results: Graphical comparison of ATE estimates within quartiles of the treatment effect

The quartiles on the x-axis are the quartiles of the treatment effect as defined by predicted CATE, the ATE per subgroup is defined by the out-of-bag CATE. The sample ATE is the difference of the average outcome between control and treatment group in the treatment effect quartile. The AIPW ATE is the augmented inverse-propensity weighted average treatment effect in the treatment effect quartile. The confidence intervals are on the 5 percent significance level. The causal forest is estimated using 0.8 of the sample as training sample with input parameters tuned once on the training sample and the domestic violence sample weight provided by DHS. The number of trees equals the number of women in the training sample (N = 37,537). The sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. Less severe violence indicates whether the woman has ever experienced at least one of the following items: pushed, shook or had something thrown; slapped; punched with fist or hit by something harmful. The outcome only refers to actions ever committed by the husband or partner.

	Quartile 1	Quartile 2	Quartile 3	Quartile 4
Quartile 1				
Quartile 2	(0.016)			
Quartile 3	0.024 (0.016)	-0.007 (0.015)		
Quartile 4	0.065 (0.016)	0.034 (0.015)	0.04 (0.014)	

Table 11: Main results: Pairwise comparison of AIPW ATE estimates within quartiles of the treatment effect

The quartiles are based on the predicted treatment effect. The difference between two quartiles is the difference in the AIPW ATE, standard errors are in parentheses. The AIPW ATE is the augmented inverse-propensity weighted average treatment effect in the treatment effect quartile. The gray background color indicates that the treatment effects are statistically significantly different from each other on the 5 percent significance level, the black background color indicates the same on the 1 percent significance level. The causal forest is estimated using 0.8 of the sample as training sample with input parameters tuned once on the training sample and the domestic violence sample weight provided by DHS. The number of trees equals the number of women in the training sample (N = 37,537). The sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. Less severe violence indicates whether the woman has ever experienced at least one of the following items: pushed, shook or had something thrown; slapped; punched with fist or hit by something harmful. The outcome only refers to actions ever committed by the husband or partner.

is statistically significantly different from zero on the 1% significance level. Between the third and the fourth quartile, the difference in the AIPW ATE is four percentage points and statistically significantly different on the 1% significance level. The difference in the AIPW ATE between the second and fourth quartile is around 3.5 percentage points and statistically significant on the 5% significance level. Only taking into account the results presented in Table 11, there are heterogeneous treatment effects among the four subgroups based on the quartiles of the predicted treatment effect.

Based on the results in the tables and figures presented above, the estimated average treatment effect is very different for the group in the fourth quartile of the predicted treatment effect compared to the other three quartiles. The average effect of having a first-born son differs between three and six percentage points between the fourth quartile and the other quartiles.

Table 12 shows the average value of the covariates in each quartiles and allows to see which women are included in the fourth group, so for which characteristics the estimated average treatment effect is different in size. Recall that the estimated average treatment effect is around minus four percentage points in the first quartile and 2.5 percentage points in the fourth quartile. In the following analysis of the results presented in Table 12, I will mostly pay attention to the difference in the covariate values between the first and fourth quartile for which the difference in AIPW ATE is largest.

Compared to the remaining quartiles, women in the fourth quartile are on average more likely to live in a wealthy household in an urban environment and they have spent on average two more years in education. Comparing the women in the first and fourth quartile, women in the first quartile are more likely to work in services and in agriculture. The average age difference to the partner is four years higher for women in the first quartile compared to the fourth quartile. At the same time the women in the first quartile are on average almost seven years younger. Almost all of the women in the first quartile were minors when having the first sexual intercourse, compared to only around half of the women in the fourth quartile. The women in the first quartile were also around three years younger at the birth of their first child.

Overall, women who are on average more likely to have experienced less severed domestic violence when having a first-born son are likely part of the upper social classes. They have a higher probability to live in wealthier household in the city, have received more education and were older at the birth of their first son. Their partners are on average only two years older and do not seem to have strong preferences for sons, given the positive treatment effect. It is unclear whether the increase in domestic violence due to the first-born son is generated by preferences for girls. Women who are on average significantly less likely to experience less severe domestic violence when having a first-born son are younger, were younger at the birth of their first child, are less wealthy and slightly more likely to live in rural parts of Colombia. They are also more likely to work in agriculture and services and to have been minors when first having sexual intercourse. Their partners are on average six years older and seem to have strong preferences for sons, given the negative treatment effect.

Table 13 shows how often each variable was used for the splitting decision relatively to the other covariates. However, this does not necessarily indicate the degree to which each of the variables is related to heterogeneity. If there are two highly correlated variables in the data set and both are related to heterogeneity to a similar extent, it is very unlikely that both will be used equally as splitting variable. Due to the random sampling composition of the training sample, the trees might split on one variable more often than on the other. If the variable that is used more often would be removed from the set of covariates, the second variable would be used as splitting variable instead. This might not affect the composition of the leaves or the predicted CATE, but it would impact the ranking in Table 13. Table 12 allows a clearer picture of which variables values vary strongly with the treatment effects.

Four variables are used in around half of the splitting decisions. The field of occupation seems to be related to heterogeneity to a smaller extent than the age difference in the relationship, the current age of the respondent, the age of the woman when giving birth to her first child and the total years spent in education. The wealth of the household and whether the women was minor at her first sexual intercourse seem to play a similar role for the heterogeneity. Overall, there is no sign of one or two variables being significantly more related to heterogeneity compared to other covariates. Instead, a set of five to six variables seems to be equally important when defining subgroups for heterogeneous treatment effects.

The four plots in Figure 8 show the predicted treatment effect across the quartiles of the four variables that were used most often to split on following Table 13 while holding the value of all other covariates constant.[17] The plots show no evidence for heterogeneity in the treatment effect for the variable values depicted. Overall, the predicted treatment effect decreases in absolute size with an increase in age and increases in absolute size with an increase of the age difference. However, none of these trends is statistically significant.

The three plots in Figure 9 show the predicted treatment effect across interactions of two variables while holding the other covariates fixed at a constant value. The interactions are the three interactions of the variable that was used most often in splitting decisions with the three following variables in Table 13. The predicted treatment effects in all plots are negative and the largest possible difference is close to 0.5 percentage points. The treatment effects are largest in absolute size for women that are in a relationship with an older man and younger than 60 percent of the women, were older when giving birth to their first child than 40 percent of the women and spent more years in education than 20 percent of the women. Overall, the probability to experience less severe violence decreases when giving birth to a first-born son when being in a relationship with a man who is relatively older than 80 percent of all men compared to their partner or wife. This can be seen as evidence for the fact that relatively

[17]It is important to note that when holding the values of all other covariates constant the predicted treatment effect for a given quartile might be averaged over only few data points, which can affect the trends plotted in the graphs.

Covariates	Quartile 1	Quartile 2	Quartile 3	Quartile 4
lives in urban environment	0.761	0.58	0.584	0.832
	(0.005)	(0.005)	(0.005)	(0.005)
wealth quintile of household	2.357	2.192	2.406	3.118
	(0.013)	(0.013)	(0.013)	(0.013)
experienced domestic violence as child	0.309	0.335	0.369	0.371
	(0.005)	(0.005)	(0.005)	(0.005)
age	30.81	33.44	37.65	37.44
	(0.081)	(0.081)	(0.081)	(0.081)
age at first child	18.07	18.9	20.96	21.74
	(0.04)	(0.04)	(0.04)	(0.04)
works in agriculture	0.022	0.103	0.089	0.009
	(0.002)	(0.002)	(0.002)	(0.002)
works in sales	0	0	0.2	0.746
	(0.003)	(0.003)	(0.003)	(0.003)
works in services	0.894	0.386	0.257	0.001
	(0.004)	(0.004)	(0.004)	(0.004)
works in skilled & manual job	0.007	0.059	0.05	0.018
	(0.002)	(0.002)	(0.002)	(0.002)
works in technical & managerial job	0.003	0.072	0.089	0.104
	(0.003)	(0.003)	(0.003)	(0.003)
works in unskilled & manual job	0.004	0.034	0.028	0.013
	(0.001)	(0.001)	(0.001)	(0.001)
total years of education	7.164	7.02	7.547	9.591
	(0.042)	(0.042)	(0.042)	(0.042)
age difference to partner	6.303	4.577	4.632	2.236
	(0.073)	(0.073)	(0.073)	(0.073)
was minor at first sexual intercourse	0.997	0.878	0.586	0.508
	(0.004)	(0.004)	(0.004)	(0.004)

Table 12: Main results: Values of covariates across the quartiles of the treatment effect

The quartiles are based on the predicted treatment effect and the values indicate the average value of the covariate in the quartile. The colors indicate the position of the mean of the subgroup in the standardized empirical distribution. The standardized distribution is colored from a scale of +/- 0.9. The causal forest is estimated using 0.8 of the sample as training sample with input parameters tuned once on the training sample and the domestic violence sample weight provided by DHS. The number of trees equals the number of women in the training sample (N = 37,537). The sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. Less severe violence indicates whether the woman has ever experienced at least one of the following items: pushed, shook or had something thrown; slapped; punched with fist or hit by something harmful. The outcome only refers to actions ever committed by the husband or partner. The variable "experienced domestic violence as child" refers to actions committed by the father against the mother during childhood. "Minor" refers to all women younger than 18 years.

Variable	Relative Importance
age difference to partner	0.12
age	0.12
age at first child	0.12
total years of education	0.12
wealth quintile of household	0.11
was minor at first sexual intercourse	0.09
works in sales	0.08
lives in urban environment	0.08
works in services	0.08
experienced domestic violence as child	0.08
works in agriculture	0.00
works in skilled & manual job	0.00
works in technical & managerial job	0.00
works in unskilled & manual job	0.00

Table 13: Main results: Sorted measure of variable importance

The variable importance indicates how often a variable was used as splitting variable. The causal forest is estimated using 0.8 of the sample as training sample with input parameters tuned once on the training sample and the domestic violence sample weight provided by DHS. The number of trees equals the number of women in the training sample (N = 37,537). The sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. Less severe violence indicates whether the woman has ever experienced at least one of the following items: pushed, shook or had something thrown; slapped; punched with fist or hit by something harmful. The outcome only refers to actions ever committed by the husband or partner. The variable "experienced domestic violence as child" refers to actions committed by the father against the mother during childhood. "Minor" refers to all women younger than 18 years.

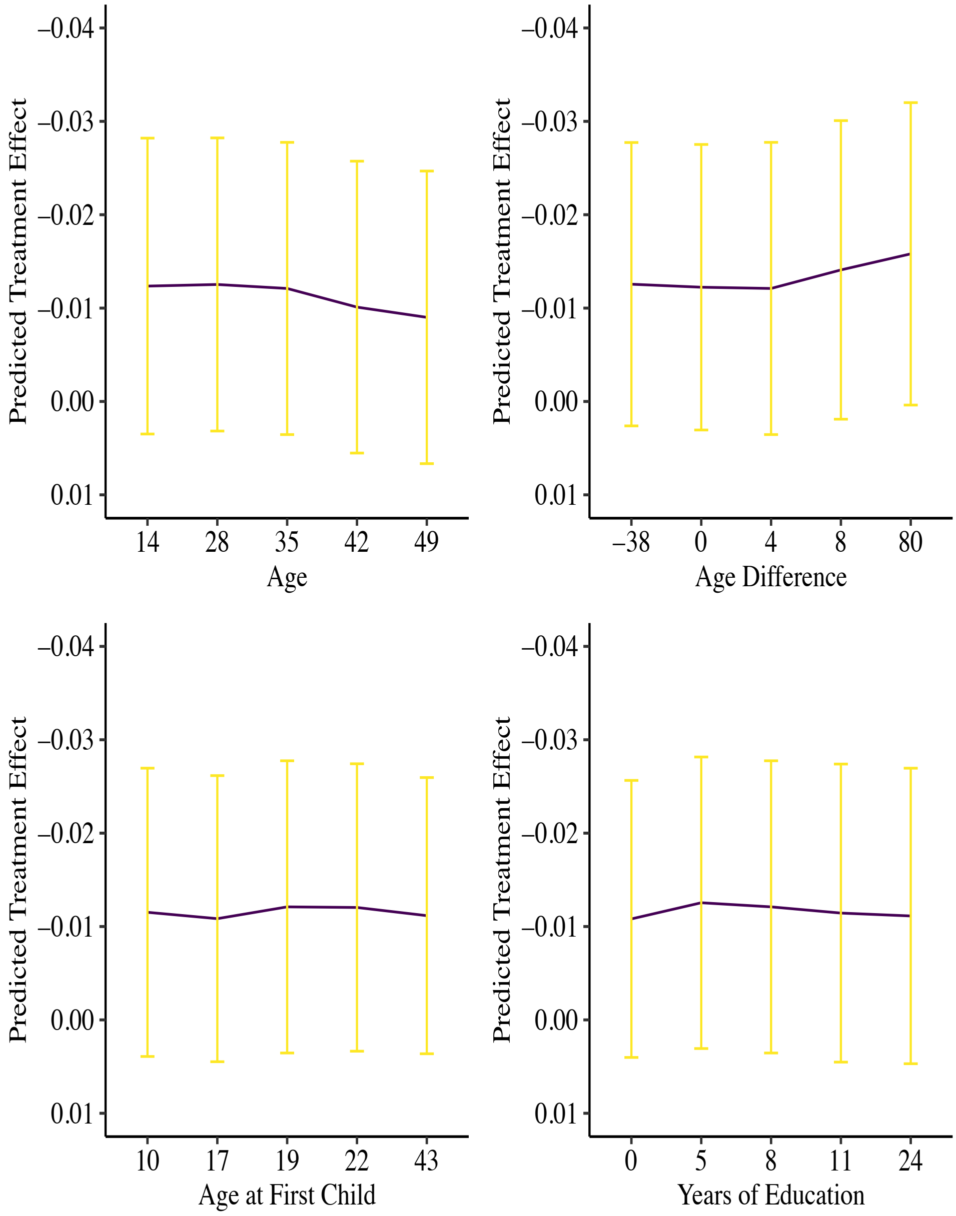

Figure 8: Main results: Partial dependency plots with a single variable

The partial dependency plots show the predicted treatment effect across the quartiles of the variable on the x-axis while holding all other covariates constant at their median value. The confidence interval is on the 5 percent significance level. The causal forest is estimated using 0.8 of the sample as training sample with input parameters tuned once on the training sample and the domestic violence sample weight provided by DHS. The number of trees equals the number of women in the training sample (N = 37,537). The sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. Less severe violence indicates whether the woman has ever experienced at least one of the following items: pushed, shook or had something thrown; slapped; punched with fist or hit by something harmful. The outcome only refers to actions ever committed by the husband or partner.

older men can be more likely to have preferences for sons and are aiming for a male heir and are at the same time also more likely to exert domestic violence when the partner gives birth to a first-born daughter.

Preliminary conclusion

When taking into account all the evidence presented in Section 6.2, it is unclear whether heterogeneous treatment effects exist for this outcome variable due to several reasons. In support of heterogeneous treatment effects are the findings based on the subgroups of the quartiles of the predicted treatment effects which show heterogeneity in average treatment effects of three to six percentage points (see Table 11). Furthermore, the subgroups defined over the covariate values as presented in Table 12 show mainly two different groups for which the average treatment effect is statistically significantly different on the 1% significance level. Women who are less likely to experience domestic violence when having a first-born son compared to a daughter are on average more likely to live in rural parts of Colombia, have a relatively older husband/partner compared to their own age, are generally younger and more likely to work in agriculture and services. Women who are more likely to experience domestic violence when having a first-born son compared to a daughter are on average older, living in cities and more likely part of the upper social classes. Findings that do not support a potential existence of heterogeneous treatment effects in this setting are the fact there are no variables that were used significantly more often for the splitting decisions compared to others, which would be a strong indicator for heterogeneity in treatment effects and the lack of significance in the partial dependency plots for the variables that were used most often for the splitting. It is important to take into account that the expected magnitude of the treatment effect is rather small, due to the vast number of potential risk factors to experience domestic violence. This makes it more difficult to assess its heterogeneity. Additionally, the outcomes are defined broadly and the timing information is scarce, which contributes to the difficulty when studying causal effects. The combination of the two last-mentioned reasons can offer an explanation why I did not find heterogeneity in treatment effects for severe and sexual domestic violence, which are less common compared to less severe domestic violence in Colombia based on Table 1. Accordingly, the lack of heterogeneity would not be a sign of no actual heterogeneity, but should be attributed to a lack in power to detect it.[18]

Additional concerns go beyond the results presented in Section 6.1 and Section 6.2 and question among others the validity of the identification strategy and the resulting selection of the sample, and more generally, the data as is. The data set consists of interviews conducted with the women, and is therefore subject to subjective statements and potential biases, such as reporting or attrition bias. Section 7 aims to provide additional information on the extent of some of these biases and the robustness of the results presented in Section 6.2. Section 8 will include a broader discussion of the validity of the results.

[18] I used a simple power calculator to get an impression of the power needed to detect heterogeneous treatment effects for severe domestic violence. If I aim to get 80% power on a 1% level and rely on the incidence rate reported in Table 3 between treatment and control group as orientation for the potential size of the heterogeneity in treatment effects, the number of observations exceeds the number of women included in the data set. Therefore, a lack in power cannot be rejected as plausible explanation for why I do not find heterogeneous treatment effects for severe and sexual domestic violence. Due to the larger incidence rate for less severe violence, lack of power is not a major concern for this outcome.

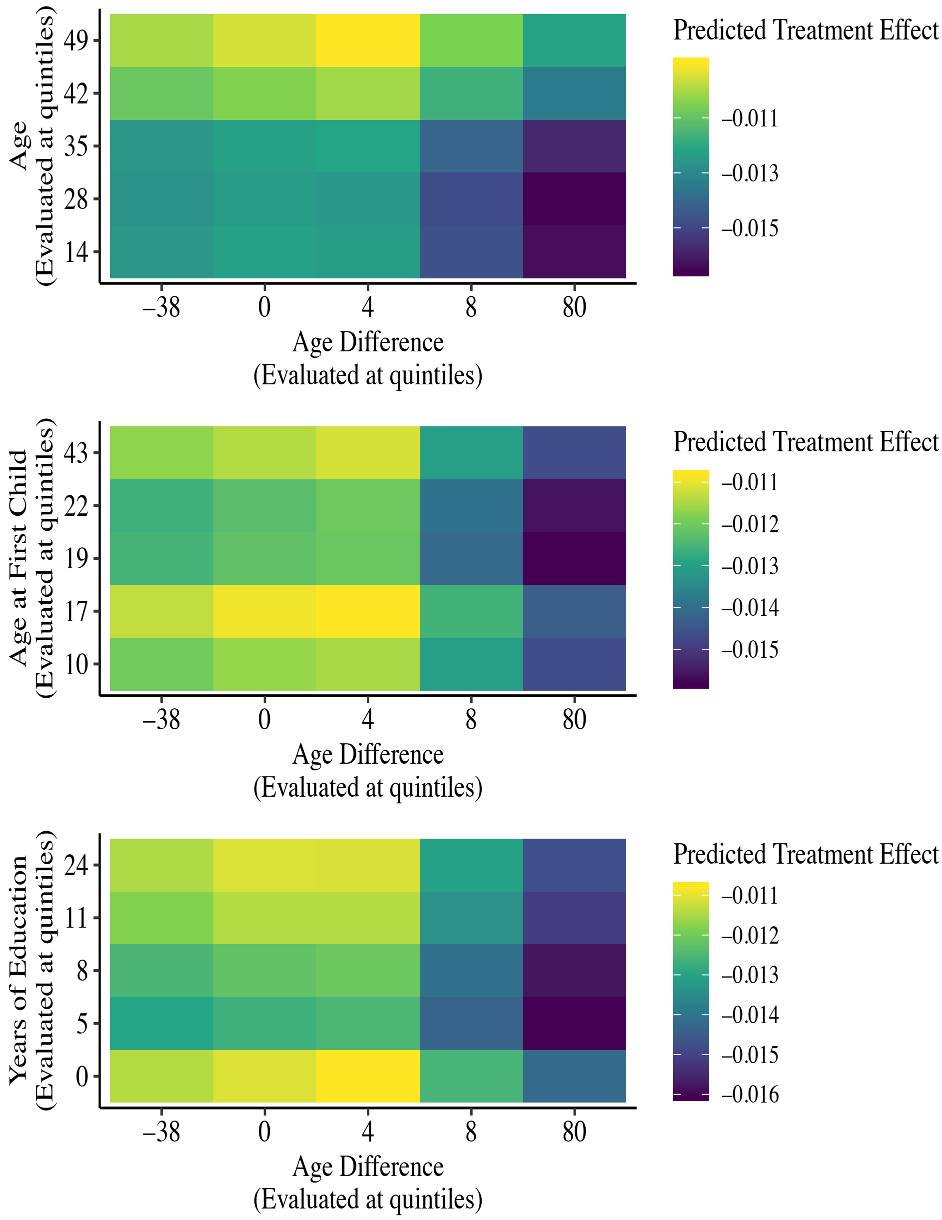

Figure 9: Main results: Partial dependency plots with two variables

The partial dependency plots show how the predicted treatment effect varies across the quartiles of the two variables while holding all other covariates constant at their median value. The causal forest is estimated using 0.8 of the sample as training sample with input parameters tuned once on the training sample and the domestic violence sample weight provided by DHS. The number of trees equals the number of women in the training sample (N = 37,537). The sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. Less severe violence indicates whether the woman has ever experienced at least one of the following items: pushed, shook or had something thrown; slapped; punched with fist or hit by something harmful. The outcome only refers to actions ever committed by the husband or partner.

7 Robustness Checks

7.1 Robustness of Heterogeneous Treatment Effects

Splitting the data set in ten subgroups

In Section 6.2 the subgroups were defined over the quartiles of the predicted treatment effect. In the following, I will split the sample into ten subgroups based on the deciles of the predicted treatment effect to decrease the number of observations per subgroup. This is supposed to provide information whether the heterogeneity in treatment effects is driven by larger or smaller subgroups and whether the difference in heterogeneous treatment effects varies when using a different number of subgroups. In Appendix Section Additional Robustness Checks, I will provide the results when defining the subgroups over the quintiles of the predicted treatment effect.

In Table 14, the sample ATEs as well as the AIPW ATEs across the deciles are statistically significantly different from zero on the 1% significance level. The AIPW ATEs are statistically significantly different from each other on the 0.1% significance level and range from minus six percentage points (third decile) to five percentage points (eight decile). The average treatment effect is statistically significantly different from zero on the 1% significance level in several of these deciles. Apparently, there is a treatment effect for a large majority of the women, but its sign and absolute size differs across the sample. The increase in the range of the values of average treatment effects per decile can be explained by the smaller sample size in each subgroup. Splitting the subgroups based on the deciles of the predicted treatment effect allows to better isolate the women who are affected by the treatment the most in absolute size and differently affected with respect to the sign of the treatment effect and provides additional evidence for heterogeneous treatment effects.

Table 15 shows which AIPW ATEs are statistically significant from each other across the deciles. The average treatment effects in the eight and ninth decile are each statistically significantly different from the average treatment effect in deciles one, two and three, as well as deciles five, six and seven on the 1% significance level. The AIPW ATE in the fourth decile is statistically significantly different from the average treatment effect in the first and third decile on the 1% significance level. The difference in average treatment effects ranges from minus 6.8 percentage points to 11.3 percentage points between the two largest average treatment effects in deciles three and eight. It is not surprising that the difference in average treatment effects is larger compared to Table 11, since splitting the subgroups into ten parts versus four allows to create subgroups with a more similar predicted treatment effect so that the average is higher for the subgroups with the largest absolute treatment effect.

Extending the main sample

A general problem when using alternative samples to re-estimate the heterogeneous treatment effects is the sample size. Setting further constraints on the main sample used throughout this thesis can reduce the sample size to an extent that heterogeneity in treatment effects cannot be detected anymore, even if there is still heterogeneity in the data. Since there is no exact sample size for which this lack of power might play an important role, interpreting the results with a sample that is significantly smaller than the main sample is difficult. Therefore, any alternative sample should have a similar or larger sample size.

I extend the sample to include all women who have ever given birth to a son instead of all women who have at least one living son and I change the treatment accordingly. Here, a women is referred to as treated if the first-born child was male, regardless of whether the first-born child is still alive. This increases the sample size by 598 women

	Sample ATE	AIPW ATE
Decile 1	-0.05398	-0.049747
	(0.015788)	(0.019178)
Decile 2	0.01124	-0.018994
	(0.015372)	(0.017079)
Decile 3	-0.053251	-0.061481
	(0.015343)	(0.018376)
Decile 4	0.005041	0.034934
	(0.015411)	(0.016481)
Decile 5	-0.0437	-0.023271
	(0.015525)	(0.016652)
Decile 6	-0.037872	-0.024175
	(0.01534)	(0.016196)
Decile 7	-0.039781	-0.033465
	(0.015609)	(0.015302)
Decile 8	-0.007013	0.051047
	(0.015742)	(0.014904)
Decile 9	-0.004166	0.04556
	(0.015988)	(0.017514)
Decile 10	-0.042622	-0.012102
	(0.015581)	(0.016262)
p-value	0.0055	0

Table 14: ATE estimates within deciles of treatment effect

The deciles on the y-axis are the deciles of the treatment effect, the ATE per subgroup is defined by the out-of-bag CATE. The sample ATE is the difference of the average outcome between control and treatment group in the treatment effect decile. The AIPW ATE is the augmented inverse-propensity weighted average treatment effect in the treatment effect decile. The p-values are testing the null hypothesis: ATE is constant across the deciles. The sample ATE uses an F-Test and the AIPW ATE uses a Wald test. The causal forest is estimated using 0.8 of the sample as training sample with input parameters tuned once on the training sample and the domestic violence sample weight provided by DHS. The number of trees equals the number of women in the training sample (N = 37,537). The sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. Less severe violence indicates whether the woman has ever experienced at least one of the following items: pushed, shook or had something thrown; slapped; punched with fist or hit by something harmful. The outcome only refers to actions ever committed by the husband or partner.

	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10
Decile 1										
Decile 2	0.031 (0.027)									
Decile 3	-0.012 (0.027)	(0.024)								
Decile 4	0.085 (0.027)	0.054 (0.024)	0.096 (0.026)							
Decile 5	0.026 (0.027)	-0.004 (0.024)	0.038 (0.026)	-0.058 (0.023)						
Decile 6	0.026 (0.027)	-0.005 (0.024)	0.037 (0.026)	-0.059 (0.023)	-0.001 (0.024)					
Decile 7	0.016 (0.027)	-0.014 (0.024)	0.028 (0.026)	-0.068 (0.023)	-0.01 (0.024)	-0.009 (0.023)				
Decile 8	0.101 (0.027)	0.07 (0.024)	0.113 (0.026)	0.016 (0.023)	0.074 (0.024)	0.075 (0.023)	0.085 (0.022)			
Decile 9	0.095 (0.027)	0.065 (0.024)	0.107 (0.026)	0.011 (0.023)	0.069 (0.024)	0.07 (0.023)	0.079 (0.022)	-0.005 (0.021)		
Decile 10	0.038 (0.027)	0.007 (0.024)	(0.026)	-0.047 (0.023)	0.011 (0.024)	0.012 (0.023)	0.021 (0.022)	-0.063 (0.021)	-0.058 (0.025)	

Table 15: Pairwise comparison of AIPW ATE estimates within deciles of the treatment effect

The deciles are based on the predicted treatment effect. The difference between two deciles is the difference in the AIPW ATE, standard errors are in parentheses. The AIPW ATE is the augmented inverse-propensity weighted average treatment effect in the treatment effect decile. The gray background color indicates that the treatment effects are statistically significantly different from each other on the 5 percent significance level, the black background color indicates the same on the 1 percent significance level. The causal forest is estimated using 0.8 of the sample as training sample with input parameters tuned once on the training sample and the domestic violence sample weight provided by DHS. The number of trees equals the number of women in the training sample (N = 37,537). The sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. Less severe violence indicates whether the woman has ever experienced at least one of the following items: pushed, shook or had something thrown; slapped; punched with fist or hit by something harmful. The outcome only refers to actions ever committed by the husband or partner.

	Sample ATE	AIPW ATE
Quartile 1	-0.024668	-0.008118
	(0.009731)	(0.010523)
Quartile 2	-0.031501	-0.034968
	(0.009705)	(0.010408)
Quartile 3	-0.015119	0.005322
	(0.009729)	(0.010198)
Quartile 4	-0.033131	0.009309
	(0.009893)	(0.010609)
p-value	0.5498	0.0117

Table 16: Extended sample: ATE estimates within quartiles of treatment effect

The quartiles on the y-axis are the quartiles of the treatment effect, the ATE per subgroup is defined by the out-of-bag CATE. The sample ATE is the difference of the average outcome between control and treatment group in the treatment effect quartile. The AIPW ATE is the augmented inverse-propensity weighted average treatment effect in the treatment effect quartile. The p-values are testing the null hypothesis: ATE is constant across the quartiles. The sample ATE uses an F-Test and the AIPW ATE uses a Wald test. The causal forest is estimated using 0.8 of the sample as training sample with input parameters tuned once on the training sample and the domestic violence sample weight provided by DHS. The number of trees equals the number of women in the training sample (N = 37,935). The sample is restricted to women who have ever given birth to at least one son, have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. Less severe violence indicates whether the woman has ever experienced at least one of the following items: pushed, shook or had something thrown; slapped; punched with fist or hit by something harmful. The outcome only refers to actions ever committed by the husband or partner.

and reduces the percentage of treated women to 66 percent. Table 16 shows the sample ATEs and AIPW ATEs for the four subgroups based on the quartiles of the predicted treatment effect. The AIPW ATEs are statistically significantly different from each other on the 2% significance level and range from minus 3.5 percentage points to essentially zero percentage points. Only the AIPW ATE in the second quartile is statistically significantly different from zero on the 1% significance level.

Table 17 shows the difference in AIPW ATEs across the four subgroups. The AIPW ATE in the second quartile is statistically significantly different on the 1% significance level from the average treatment effect in the third and fourth quartile, the differences are four and 4.4 percentage points. Overall, Table 16 and Table 17 suggest that the average treatment effect in the second subgroup is different from the remainder of the sample. Since the absolute size of the average treatment effect is smaller and there are no statistically significant positive treatment effects, including women who have experienced a loss of their son reduces the absolute size of treatment effect. This indicates that these women are less likely to experience a reduction in less severe domestic violence due to the fact that their first-born child was male. This could be attributed to the fact that the partner or husband cares more about a living son or the loss of the child generally leads to an increase in domestic violence even though the gender of the child might have lead to a reduction in the domestic violence behavior.

7.2 Sample Bias

The goal of this subsection is to provide more information on potential bias in the sample due to its construction. In the main sample, I oversample women with only one child, a son, who are likely to be younger, still in the family planning process and to have been married for a period below average. Women with more children are likely to have been exposed longer to their partner or husband. If there is an exposure effect, the oversampled women are

	Quartile 1	Quartile 2	Quartile 3	Quartile 4
Quartile 1				
Quartile 2	(0.015)			
Quartile 3	0.013 (0.015)	0.04 (0.015)		
Quartile 4	0.017 (0.015)	0.044 (0.015)	0.004 (0.014)	

Table 17: Extended sample: Pairwise comparison of AIPW ATE estimates within quartiles of the treatment effect

The quartiles are based on the predicted treatment effect. The difference between two quartiles is the difference in the AIPW ATE, standard errors are in parentheses. The AIPW ATE is the augmented inverse-propensity weighted average treatment effect in the treatment effect quartile. The gray background color indicates that the treatment effects are statistically significantly different from each other on the 5 percent significance level, the black background color indicates the same on the 1 percent significance level. The causal forest is estimated using 0.8 of the sample as training sample with input parameters tuned once on the training sample and the domestic violence sample weight provided by DHS. The number of trees equals the number of women in the training sample (N = 37,935). The sample is restricted to women who have ever given birth to at least one son, have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. Less severe violence indicates whether the woman has ever experienced at least one of the following items: pushed, shook or had something thrown; slapped; punched with fist or hit by something harmful. The outcome only refers to actions ever committed by the husband or partner.

more likely to have experienced less domestic violence due to the lower exposure and not the gender of their only child.

In the following, I will try to approximate the exposure effect by looking at the number of the children, the number of sons and the age of the women as proxies for exposure.

Does the number of children or the number of sons affect the probability of having experienced domestic violence?

As seen in Figure 10, the percentage of women who experienced less severe domestic violence varies less than 10 percentage points with the number of children. Women with six or seven children are most likely to have experienced domestic violence, whereas women with only one child or more than eight children are the least likely. Similarly, the percentage of women who experienced less severe domestic violence varies less than ten percentage points with the number of living sons over the number of living daughters. Based on the number of children per gender, for each fixed number of living daughters up to four, the probability of having experienced domestic violence increases up to the point of five or six children in total and decreases afterwards. When holding the number of living sons constant, we see that among mothers who have one living son, mothers with between two and four daughters have the highest probability of having experienced domestic violence. The difference in probability decreases with additional sons, until all mothers report around the same degree of domestic violence regardless of the number of living daughters when they have three to four sons. For women with five or more sons, the patterns are reversed and women with one or five daughters report the most domestic violence.

If we interpret the number of children as a perfect proxy for exposure, since women who have more children have likely been exposed for a longer time to their husband or partner, and we anticipate an exposure effect, we would expect to see a strict increase in the ratio of women who have experienced domestic violence with the number of children. However, this is not the case, because the number of children also reflects other factors, such as potentially more stress or more frustration over the gender of the children. Interestingly, the graphs only show the fact whether a woman ever experienced a certain type of domestic violence and we also know that a woman with, e.g. six children went through the stages of having one child, two children and so on. The fact that women with more children are less likely to report domestic violence can hint towards two major reasons: Women who have more children have spend more time with their partner and are more reluctant to report domestic violence or are less likely to recall domestic violence that occurred a longer time ago when they had less children. Alternatively, women with more children are actually less likely to have experienced domestic violence, which seems an interesting phenomenon. Consequently, families that have more children and have always planned to have more children would be systematically less likely to experience domestic violence, which would then be correlated to characteristics that impact the number of children. One reason could be that families with more children are more likely to have a son and therefore, women experience less domestic violence due to unfulfilled preferences. Contrary, the frustration over not having a son increases with the number of daughters up to a certain point. The graph looking at the number of children per gender hints towards a certain number of children, where women with the corresponding number of children are less likely to report domestic violence compared to women with less children.

In terms of the exposure effect, the following can be concluded: the number of children is, as expected, not a perfect proxy for exposure effects. The explanations given for the graphs are mainly either a difference in reporting behavior for women with more than five children or structural differences in larger families compared to smaller ones that lead to less domestic violence in the household. There is no clear evidence for a strong exposure effect, the patterns can instead be partly explained by different reasons. This could be supported if women are most likely to report domestic violence when they are younger and at the beginning of their relationship, which is found by, among others, the DHS (Kishor and Kiersten, 2004).

In which period of their life are women reporting the highest probability of experiencing domestic violence?

Figure 11 shows the reported less severe domestic violence ever and during the last year, conditional on the age of the woman. We see that with an increase in age, women are less likely to report domestic violence, which could either be due to a difference in reporting behavior of experienced domestic violence, such as not remembering the incident or an unwillingness to report it or a generational gap, such that older women were actually less likely to have experience domestic violence. The reported domestic violence ranges from around 35 percent to 30 percent, a difference of five percentage points. For the reported less severe domestic violence during the last year, we see that women are most likely to have been domestically abused when they are young and potentially at the beginning of the relationship. The reported domestic violence ranges from around 30 percent to less than 15 percent, a difference of more than 15 percentage points. Comparing both graphs, we see that older women are less likely to have experienced domestic violence during the last year, even if they would be more reluctant to report it if we the difference in the reported violence with respect to the time period as evidence for this behavior. Therefore, oversampling young women is likely to lead to over-proportional reports of domestic violence during the last year, but might still be accurate for the domestic violence experienced in their lifetime, since women are most likely to be domestically abused in the first years of their relationship.

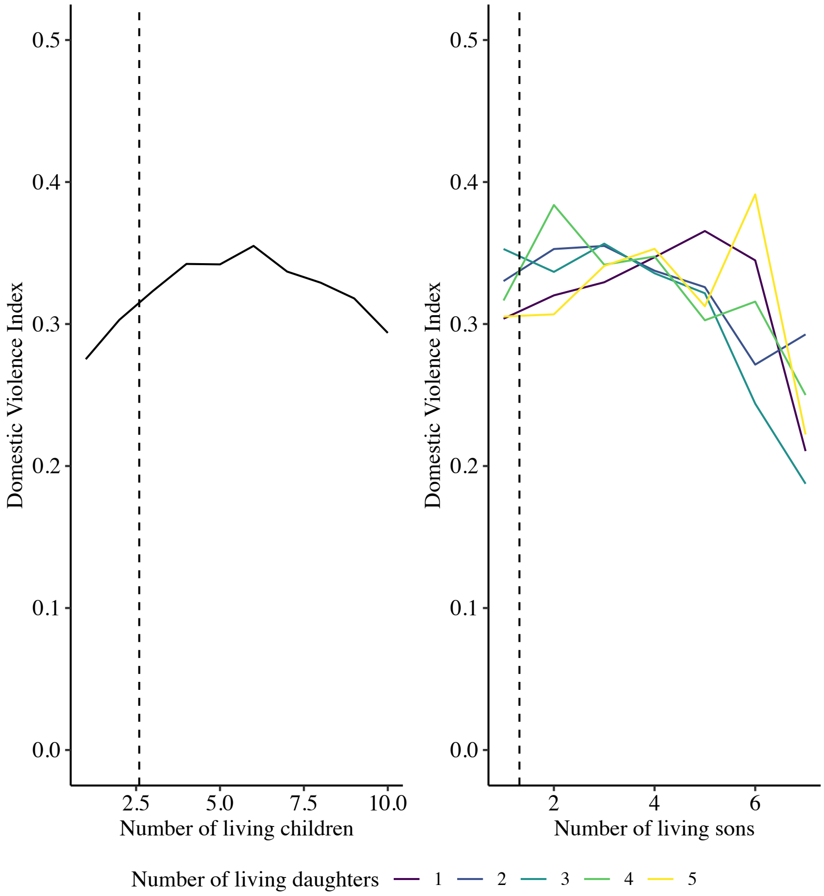

Figure 10: Sample bias: Number of living children and living sons against experienced domestic violence

The dashed line indicates the mean of the variable on the x-axis. The sample is restricted to women who have at least one living child and non-missing values for the outcomes, the age, the number of living children and living sons (N = 78,194). In the right panel, the sample is additionally restricted to women with at least one living son. For the number of children, the x-axis is restricted to a sample size of more than 100 women for each number of children. For the number of living sons, the number of children for each gender is restricted to 5 in order to avoid small sample sizes and extreme values of experienced domestic violence. The domestic violence index indicates whether the woman has ever experienced at least one of the following items: pushed, shook or had something thrown; slapped; punched with fist or hit by something harmful. The outcome only refers to actions ever committed by the husband or partner.

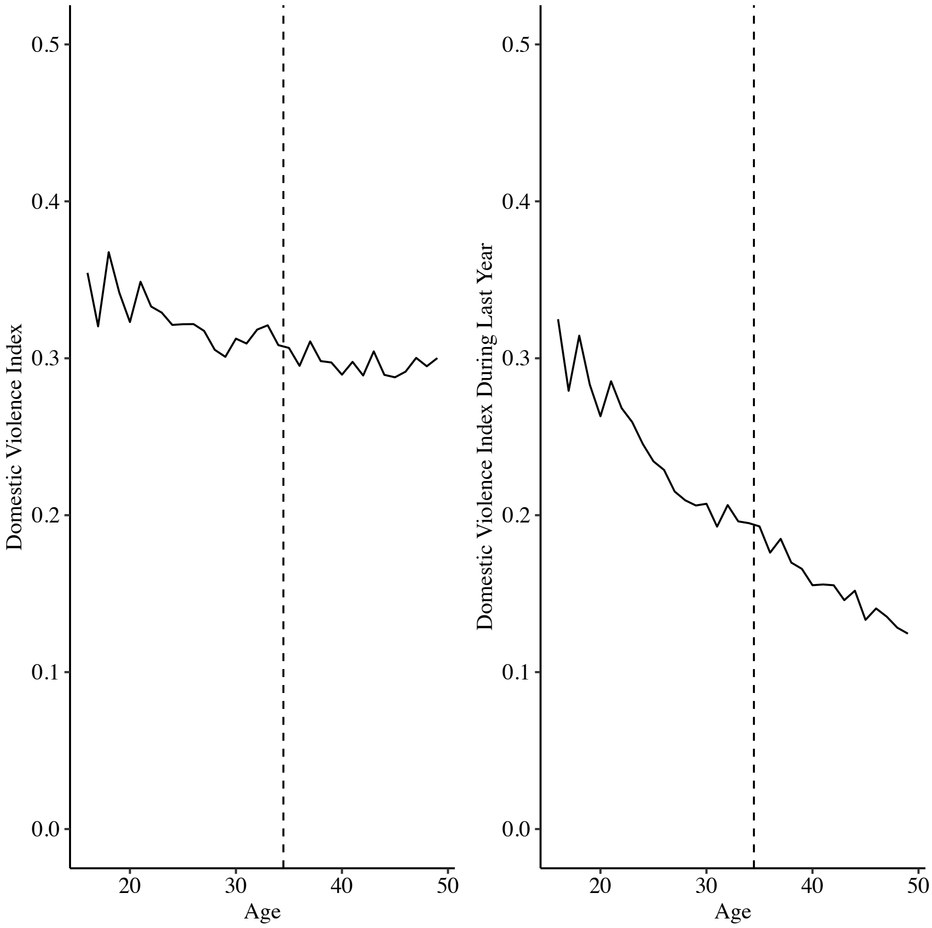

Figure 11: Sample bias: Age of women against experienced domestic violence

The dashed line indicates the mean age in the sample. The sample is restricted to women who have at least one living child and non-missing values for the outcomes, the age, the number of living children and living sons (N = 78,194). The x-axis is restricted to a sample size of more than 100 women with the corresponding age. The domestic violence index indicates whether the woman has experienced at least one of the following items during the last year: pushed, shook or had something thrown; slapped; punched with fist or hit by something harmful. The outcome only refers to actions ever committed by the husband or partner in the left panel and actions committed during the last year by the husband or partner in the right panel.

7.3 Reporting and Attrition Bias

Reporting bias

Since I only have access to this data set, it is impossible to detect systematic reporting bias in the given setting. However, there are some smaller potential differences in reporting behavior, which are based on observables and can be analyzed. Women who have split from their partner have a lower threshold to report domestic violence compared to women who are still married to the abusive partner (González-Brenes, 2004). At the same time, they are more likely to have experienced more severe domestic violence more frequently, which was potentially among the reasons for the separation. I can compare reports of experienced domestic violence of women to check for different reporting behavior depending on the current relationship with the abusive partner.

Figure 12 compares the reports of less severe domestic violence during and before last year for women with different current relationship stati. We see that women who are living with their partner or are still married to the partner are less likely to report domestic violence compared to women who are currently separated, divorced or widowed. Married women are likely to be older compared to women who are currently living with their partner and are therefore more likely to have experienced domestic violence before the last year rather than during the last year prior to the interview. Also, widowed women are less likely to have experienced domestic violence last year compared to the year before last year prior to the interview. Contrary, separated women report more domestic violence during the last year, which could be the reason for their separation, which has not ended in a divorce yet. Overall, the patterns themselves and the patterns in comparison to patterns of other relationship stati can be explained by the relationship status. Even though this no clear evidence for or against differences in reporting behavior, the last-mentioned is also no apparent cause to worry. For example, widowed women reporting more domestic violence during the last year compared to the year before last year would threaten the credibility of the reported domestic violence.

Attrition bias

In the main data set, I exclude women who have missing values for the outcomes and the main covariates. As demonstrated in Table 19 in the Appendix, there are in total 48,343 women in the entire sample of around 133,583 who have missing outcomes and 72,253 in the entire sample with at least one missing value for one of the covariates. However, not all of these women would have been included in the sample had they had non-missing values for either the outcomes or the main covariates due to the additional sample constraints. Therefore, I start with the original data set of 133,583 women and exclude all women who do not have a living son, which is non-missing for all women and which reduces the sample to 65,188 women.

Table 18 shows the missings of each variable in the reduced data set.

It is possible to remove all missings in the data set in three steps:

1. Remove all missings for the variable “age difference in the relationship”
2. Remove all remaining missings for the outcomes
3. Remove all remaining missings for the variable “was minor at the first sexual intercourse”

In order to identify potential attrition bias, I compare all individuals that are removed in each step from the data set with the individuals that are in the main sample based on the means of the remaining covariates to get an idea

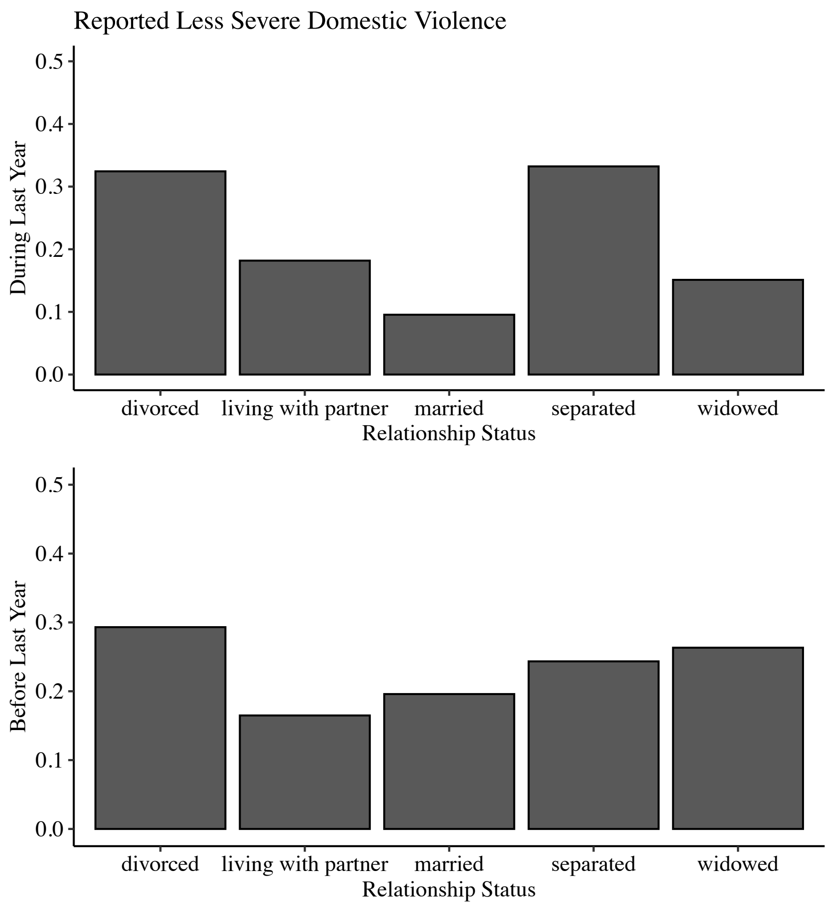

Figure 12: Reporting bias: Reporting behavior among women with different relationship stati

The x-axis shows the mean of the women who have experienced less severe domestic violence during or before last year, grouped by relationship status. The sample contains all women from the original sample (N = 133,583) with non-missing values for the outcomes, which reduces the sample used in the graph to N = 85,240. In this sample, 447 women are divorced, 16,794 are separated, 1,998 are widowed, 43,710 are living together with their partner and 22,291 are married. It is important to note that the missing outcomes are not proportional to the group sizes. 511 married women, 12 divorced, 207 separated, 29 widowed and 791 women living together with their partner refused to answer the question. Less severe violence indicates whether the woman has experienced at least one of the following items: pushed, shook or had something thrown; slapped; punched with fist or hit by something harmful. The outcome only refers to actions committed by the husband or partner.

Variable	Missings
ever experienced less severe domestic violence	5389
ever experienced severe domestic violence	5389
ever experienced sexual domestic violence	5389
experienced domestic violence as child	2426
lives in urban environment	0
wealth quintile of household	0
age difference to partner	14655
age	0
age at first child	0
was minor at first sexual intercourse	2165
has ever been in a union, marriage or relationship	4236
works in agriculture	0
works in clerus	0
works in sales	0
works in services	0
works in skilled & manual job	0
works in technical & managerial job	0
works in unskilled & manual job	0
total years of education	0
treatment	0

Table 18: Missings in sample of all women with one living son

The table shows the missings of each variable that is used to define the main sample used in the thesis. The original data set contains N = 133,583 observations and the sample of all women with at least one living son N = 65,188 observations. The variable "experienced domestic violence as child" refers to actions committed by the father against the mother during childhood. "Minor" refers to all women younger than 18 years.

whether these women are different from women in the main sample. The corresponding three tables, Table 36, Table 37 and Table 38 can be found in Appendix Section Additional Robustness Checks. It is obvious that this exercise cannot capture the entire extent of the attrition bias, but it hopefully gives an impression of its severity. There are two things to keep in mind with respect to the differences shown in these tables. First, some women that have missings in the respective variables also have missings for some of the other covariates. Therefore, the means in the right column of each table are calculated over a subset of the women with missings in the respective variable mentioned in the enumeration above. However, some variables in Table 18 do not have any missings and should allow a clearer picture. Second, some of the differences in the covariates are, even though they are statistically significant, very small and it is unclear whether these differences are random or should actually be used to explain differences between the two groups. Third, the table does not show any causal relations, so it is only likely that the covariates that are statistically significantly different in the two groups are impacting the probability of providing the specific information.

The women with missings in the respective variables are generally different in many covariates compared to the main sample. Women eliminated in Step 1, who have a missing value for the age difference, are more likely to have a first-born son and are also more likely to have experienced any form of domestic violence. This is problematic, given that the current results suggest that treated women are less likely to experience domestic violence and the difference in age is one of the main identified drivers as seen in Table 13 and Table 33. This effect could be caused by the women who are not indicating the age of their partner and who are more likely to be treated and to experience domestic violence. Women eliminated in Step 2 who did not indicate whether they had ever experienced domestic violence, but indicated a value for the age difference are more likely to be treated, to live in urban environments and in wealthier households. They are on average younger, were older at the birth of their first child and received more education compared to women in the main sample. Women eliminated in Step 3 who indicated the outcomes and the age difference, but not whether their were minors at their first sexual intercourse are not more or less likely to be treated, but are less likely to have experienced domestic violence as a child, are less likely to live in urban places and live in less wealthy households.

Overall, these findings question the validity of the results. Also taking into account the sample sizes of the women eliminated in each step, the first step shows to be the most concerning one. Women who do not indicate the age of their partner make up around 25% of all women in the main sample and are more likely to be treated and to have experienced domestic violence.

In order to get an approximation for the effect this might have on the results, I predict the value of the age difference in the relationship using a tuned regression forest for the 10,770 women for which this is the only missing value among the variables in Table 18 and rerun the causal forest for the outcome whether the women ever experienced less severe domestic violence. Table 39, Table 40, Table 41 and Table 42 in Appendix Section Additional Robustness Checks show the results of the causal forest.

As demonstrated in Table 39, the AIPW ATEs across the subgroups based on the quartiles of the predicted treatment effect are statistically significantly different on the 4% significance level and range from close to zero percentage points to minus three percentage points, whereby the last-mentioned is statistically significantly different from zero on the 1% significance level. The average treatment effect of minus three percentage points is statistically significantly different from the average treatment effects in the third and fourth quartile on the 5% significance level, the difference is around three percentage points, based on Table 40. The differences in the covariate values across the subgroups in Table 41 are very similar to the ones described in Table 12. The age difference in the relationship was used five percentage points more often for the splitting decision in Table 42 compared to the main results in Table 13. This is a difference of two percentage points to the following variable, the age at the first child, which supports the previous evidence that the age difference is one of the closest related variables to the heterogeneity in treatment effects. Overall, the heterogeneity decreases for the negative treatment effect and the positive treatment effect is not statistically significantly different from zero anymore. However, compared to the results in Section 6.2 the difference can be attributed to several reasons, only one of which is the reduction in attrition bias. Other potential explanations are, among others, the use of the remaining covariates to predict the relationship age difference, which also leads to less extreme predictions that are closer to the mean compared to the distribution of the age difference in the main sample and the increase in sample size. Overall, this exercise does not fully support the results presented in Section 6.2 and might reduce, but should not eliminate the concerns of attrition bias. The previous results seem to hold, even though to a lower extent, when increasing the sample size to include women with missing values for only the age difference in the relationship, which was previously identified as one of the main drivers in heterogeneity.

8 Discussion and Conclusion

This thesis aimed to identify subgroups of women who experienced different reactions of their husbands/partners in terms of domestic violence to the first child being a boy instead of a girl. I find that there are mainly two different subgroups of women, for which the absolute size of the treatment effect differs around six percentage points. Women who are less likely to experience domestic violence when having a first-born son compared to a daughter are on average younger, have a larger age gap to their partner and are more likely to live in rural areas. On the opposite, women who are apparently more likely to experience domestic violence when having a first-born son compared to a daughter are on average older, are living in cities and are more likely part of the upper social classes. This thesis makes contributions to several fields in Economics. First, I study a risk factor of domestic violence that has not been, to the best of my knowledge when I started working on this thesis, the main focus of previous research, nor evaluated in-depth before. Second, I use non-standard Machine Learning methods to identify heterogeneous treatment effects over observable characteristics of the women. The causal forest algorithm used in this thesis is data-driven and is suited to estimate causal treatment effects. The generated output allows to get an impression of potential subgroups for which the treatment effect is statistically significantly different. This approach has several advantages to common methods for the identification of heterogeneous treatment effects, since these methods can be problematic or even compromise the validity of the results. Furthermore, the estimation procedure of the causal forest is doubly-robust and therefore takes into account the difference in propensity score for the individual women, an issue caused by the identification strategy used in this thesis.

The main assumption on which the identification strategy heavily relies is the exogeneity of the treatment, the gender of the first child. I present general evidence for this assumption in the literature overview in Section 2 and specific evidence on the exogeneity in the data set in Subsection 4.3. However, even if this assumption is likely fulfilled, additional issues have to be considered when interpreting the results and which might even threaten their validity.

First, I do not find one variable or a small set of variables that were used significantly more often for the splitting decisions in the causal forest compared to the remaining variables. Furthermore, the partial dependency plots presented in Section 6.2 show a lack of significant variation in the size of the treatment effect over the quartiles of the tested variables. Both of these would be strong indicators for heterogeneity in treatment effects, even though their absence does not equal the absence of heterogeneous treatment effects in this setting. Additionally, the findings presented in Section 7 are not strong enough to eliminate concerns of sample bias in (un-)observable characteristics due to the identification strategy and concerns of attrition bias. I find that women with missings are likely different to those without missings in the answers they did provide in the interviews. When I attempt to predict the value of the most commonly missing variable and rerun the causal forest algorithm, the heterogeneity in treatment effects is smaller and less significant.

Second, I oversample women who only had one child and are still in the family planning stage in the treatment group compared to the control group and I undersample these women in the main selected sample compared to all women in the original data set which is caused by the identification strategy presented in Section 3. I find that women with at least one son and their families are different from those with at least one child in Section 4.4 and that women in the treatment group differ from those in the control group in a few observable characteristics in Section 4.3. This results in different estimated distributions of the propensity score for women with at least one child/at least one son and women in the control and treatment group as demonstrated in Section 4.4. This is not an immediate concern if the propensity score is estimated correctly, since the causal forest is doubly-robust. However, I only used a limited set of variables to do so, which likely results in an incorrect estimation of the propensity score. Similarly, the limited set of variables incorrectly specifies the conditional mean function. Taken together, the estimated treatment effects are likely biased due to omitted variables, even though the estimates are more robust compared to common methods such as the linear probability model.

Third, the data set provided by the DHS for Colombia is not perfectly suited to study the research question, which is due to a lack in timing information, a broad definition of the outcomes and the fact that the data is gathered through interviews and is therefore self-reported. The ideal setup to study the research question is a longitudinal study, in which women are followed over a longer time period and each incident of domestic violence they experience is recorded. This would allow to exactly determine differences in domestic violence experienced before and after the birth of the first-born child, which would subsequently allow to compare these differences over women, conditional on the gender of the first child. For obvious reasons, this setup is extremely difficult to obtain in reality, if not impossible. However, the current data set and the interviews could be improved with respect to the research question of this thesis. Currently, the only information on the timing of the experienced domestic violence is whether it happened ever, before or during the last year. Additional time horizons would allow to better match the experienced domestic violence to the birth of the first child and therefore, to estimate the treatment effect more precisely. It should be noted at this point that data sets containing information on domestic violence are very scarce and the one I use is likely among the best sources available to study the research question of this thesis.

As demonstrated above, the results should not be interpreted without some degree of caution with respect to the internal validity given the robustness checks, the estimation procedure and the data set which is not perfectly suited to answer the research question of the thesis. But even if all of these concerns were eliminated, the policy recommendations of the results are unclear.

The estimated treatment effects range from minus four to positive 2.5 percentage points in the main specification using four different subgroups and the resulting difference of six percentage points or more than a fifth of the rate of experienced domestic violence among all women in the data set, based on Table 1, is statistically significantly different from zero on the 1% significance level. For comparisons, Angelucci (2008) evaluates a CCT in Mexico and finds reduction rates of experienced domestic violence of 1.6 to 3.4 percentage points. Bobonis et al. (2013) analyze the same CCT program and estimates reductions of five to seven percentage points. Hidrobo and Fernald (2013) studies a CCT program in Ecuador and identifies a decrease in controlling behavior of six percentage points. Taking into account that I observe the treatment effect on ever experienced less severe domestic violence, the impact of the gender of the first child on domestic violence seems to be quite substantial and similar in size to the ones of the CCTs listed above. Obviously, potential policy implications differ dramatically, since the gender of the first child is exogenous, whereas CCTs can be implemented.

The finding of heterogeneous treatment effects results in different questions which have to be addressed prior to implementing or changing policies in a way that reflect the results. In the following, I will highlight some of the potential concerns related to heterogeneous treatment effects. To simplify the reasoning presented below, I will assume that there is a fixed budget for supporting women who experienced domestic violence. Naively, one is tempted to interpret the results as an imperative to act and to allocate some of the budget towards the women who are most at risk to experience an increase in domestic violence due to the gender of the first child. The support could, for example, target women who give birth in hospitals. If the characteristics of the woman fit the profile of one of the two subgroups at risk and the woman gives birth to a child with the corresponding gender, immediate actions could be prompted such as talking to her husband/partner or offering information on support and legal rights. However, as stated above, this is a naive approach and should not be practiced in reality without further information. Instead, the following considerations have to be taken into account:

First, the overall question is whether it is more useful to re-allocate the budget towards targeting the identified risk factor instead of providing support that aims at other risk factors. The largest effect of the gender of the first child on domestic violence in this paper appeared when dividing the sample in ten subgroups and ranged from minus six to plus five percentage points. Based on these numbers, women with the same characteristics as the ones in the subgroup, but with a child of the different gender will therefore on average experience an increase in domestic violence of around five to six percentage points. However, addressing other risk factors could lead to more substantial reductions in experienced domestic violence for a larger subset of women. If we want to decrease the overall rate of experienced domestic violence given the fixed budget, it would be more useful in this setting to focus on these other risk factors instead of supporting the subgroups of women described above whose first child has the gender that puts them at risk of experiencing an increase in domestic violence. It is obvious that the costs of reduction for each risk factor have to be taken into account. Preferably, the focus should be on the most cost-efficient method to achieve the largest reduction in domestic violence attributable to the corresponding risk factor. How to do so opens up an entirely new set of questions, since this information is currently not available.

Second, if there is an agreement to address the newly identified risk factor, it is unclear whether the support should focus on women with specific characteristics, which have to be defined based on the output of the causal forest instead of a broader sample, including potentially even all women. Currently, there is no clear measure or guideline on how to interpret the output of the causal forest algorithm, which makes it difficult to clearly identify the characteristics of the subgroups that are expected to experience the largest increase in domestic violence if their first child has the corresponding gender. As demonstrated in the robustness checks in Section 7, splitting the sample in a larger number of subgroups always increases the heterogeneity in treatment effects. There is, to the best of my knowledge, no research on how strong the heterogeneity in treatment effects has to be to justify addressing only subsets of individuals instead of the entire sample. Given such a threshold, the recommendation would then be to choose the number of subgroups that results in heterogeneous treatment effects with the thresholded difference.

This procedure still leaves the question on how to describe the subgroups in terms of observable characteristics, since most likely not all observed differences in characteristics can be attributed to the heterogeneity in treatment effects. Focusing on the variables that were used most often for the splitting decisions is not uncritical, since there are many different ways to describe the same subgroups. If two important variables are highly correlated, this can result in these two variables being used moderately for the splitting decisions instead of one of them being used extensively if it was the only one of the two included in the data set. Since we currently neither have a threshold that allows to justify "enough" heterogeneity nor the means to correctly identify the relevant combination of characteristics of the subgroups that drive the heterogeneity, any action plan relies on arbitrary results. This is not only critical from a statistical point of view, but also from a moral and ethical perspective. How can we justify to spend a certain amount of the budget on specific subgroups if we only have a vague justification and a vague idea on how to define the subgroups? In the context of domestic violence, any wrong recommendation based on unclear results will negatively affect the life of many women who should have gained from the recommendation.

This thesis aims to make a small contribution to the field of domestic violence and to spread awareness of the issues surrounding it. It is obvious that further research is needed, not only to provide additional evidence surrounding the effect of the children's gender on domestic violence, but on risk factors in general and the most efficient methods to support the affected women. The overall goal is to prevent all forms of domestic violence - a goal that will most likely never be reached. But given the immense costs associated with domestic violence, its cruel and perversive nature and over 2.3 billion survivors in the world we should never stop trying.

References

Abrevaya, J. (2009). Are There Missing Girls in the United States? Evidence from Birth Data, *American Economic Journal: Applied Economics* **1**(2): 1–34.
URL: *http://pubs.aeaweb.org/doi/10.1257/app.1.2.1*

Ackerman, P. L. and Beier, M. E. (2003). Intelligence, Personality, and Interests in the Career Choice Process, *Journal of Career Assessment* **11**(2): 205–218.
URL: *http://journals.sagepub.com/doi/10.1177/1069072703011002006*

Aizer, A. (2010). The Gender Wage Gap and Domestic Violence, *American Economic Review* **100**(4): 1847–1859.
URL: *http://pubs.aeaweb.org/doi/10.1257/aer.100.4.1847*

Angelucci, M. (2008). Love on the Rocks: Domestic Violence and Alcohol Abuse in Rural Mexico, *The B.E. Journal of Economic Analysis & Policy* .
URL: *https://www.degruyter.com/view/j/bejeap.2008.8.1/bejeap.2008.8.1.1766/bejeap.2008.8.1.1766.xml*

Athey, S. and Imbens, G. (2016). Recursive Partitioning for Heterogeneous Causal Effects, *Proceedings of the National Academy of Sciences* **113**(27): 7353–7360.
URL: *http://arxiv.org/abs/1504.01132 http://www.pnas.org/lookup/doi/10.1073/pnas.1510489113*

Athey, S., Imbens, G. W. and Wager, S. (2018). Approximate Residual Balancing: Debiased Inference of Average Treatment Effects in High Dimensions, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **80**(4): 597–623.
URL: *http://arxiv.org/abs/1604.07125 http://doi.wiley.com/10.1111/rssb.12268*

Athey, S., Tibshirani, J. and Wager, S. (2019). Generalized Random Forests, *The Annals of Statistics* **47**(2): 1148–1178.
URL: *http://arxiv.org/abs/1610.01271 https://projecteuclid.org/euclid.aos/1547197251*

Athey, S. and Wager, S. (2019). Estimating Treatment Effects with Causal Forests: An Application.
URL: *http://arxiv.org/abs/1902.07409*

Barcellos, S. H., Carvalho, L. S. and Lleras-Muney, A. (2014). Child Gender and Parental Investments In India: Are Boys and Girls Treated Differently?, *American Economic Journal: Applied Economics* **6**(1): 157–189.
URL: *http://pubs.aeaweb.org/doi/10.1257/app.6.1.157*

Bharadwaj, P., Dahl, G. B. and Sheth, K. (2014). Gender Discrimination in the Family, *The Economics of the Family: How the Household Affects Markets and Economic Growth* **2**.

Bobonis, G. J., González-Brenes, M. and Castro, R. (2013). Public Transfers and Domestic Violence: The Roles of Private Information and Spousal Control, *American Economic Journal: Economic Policy* **5**(1): 179–205.
URL: *http://pubs.aeaweb.org/doi/10.1257/pol.5.1.179*

Chin, Y.-M. (2012). Male Backlash, Bargaining, or Exposure Reduction? Women's Working Status and Physical Spousal Violence in India, *Journal of Population Economics* **25**(1): 175–200.
URL: *http://link.springer.com/10.1007/s00148-011-0382-8*

Choi, E. J. and Hwang, J. (2015). Child Gender and Parental Inputs: No More Son Preference in Korea?, *American Economic Review* **105**(5): 638–643.
URL: *http://pubs.aeaweb.org/doi/10.1257/aer.p20151118*

Dahl, G. and Moretti, E. (2004). The Demand for Sons: Evidence from Divorce, Fertility, and Shotgun Marriage, *Technical report*, National Bureau of Economic Research, Cambridge, MA.
URL: *http://www.nber.org/papers/w10281.pdf*

Erten, B. and Keskin, P. (2018). For Better or for Worse?: Education and the Prevalence of Domestic Violence in Turkey, *American Economic Journal: Applied Economics* **10**(1): 64–105.
URL: *http://pubs.aeaweb.org/doi/10.1257/app.20160278*

Finnoff, K. (2012). Intimate Partner Violence, Female Employment, and Male Backlash in Rwanda, *The Economics of Peace and Security Journal* **7**(2).
URL: *http://www.epsjournal.org.uk/index.php/EPSJ/article/view/139*

Flake, D. F. and Forste, R. (2006). Fighting Families: Family Characteristics Associated with Domestic Violence in Five Latin American Countries, *Journal of Family Violence* **21**(1): 19–29.
URL: *http://link.springer.com/10.1007/s10896-005-9002-2*

Garcia-Moreno, C., Jansen, H. A., Ellsberg, M., Heise, L. and Watts, C. (2005). WHO Multi-Country Study on Women's Health and Domestic Violence Against Women, *Geneva: World Health Organization* **204**: 1–18.

Garcia-Moreno, C., Pallitto, C., Devries, K., Stöckl, H., Watts, C. and Abrahams, N. (2013). *Global and Regional Estimates of Violence Against Women: Prevalence and Health Effects of Intimate Partner Violence and Non-Partner Sexual Violence*, World Health Organization.

Godoy, R., Reyes-García, V., McDade, T., Tanner, S., Leonard, W. R., Huanca, T., Vadez, V. and Patel, K. (2006). Why Do Mothers Favor Girls and Fathers Boys?, *Human Nature* **17**(2): 169–189.
URL: *https://doi.org/10.1007/s12110-006-1016-9*

González-Brenes, M. (2003). Domestic Violence, Bargaining and Fertility in Rural Tanzania.

González-Brenes, M. (2004). Domestic violence and household decision-making: Evidence from East Africa.

Hainmueller, J. (2012). Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies, *Political Analysis* **20**(1): 25–46.
URL: *https://www.cambridge.org/core/product/identifier/S1047198700012997/type/journal_article*

Heath, R. (2014). Women's Access to Labor Market Opportunities, Control of Household Resources, and Domestic Violence: Evidence from Bangladesh, *World Development* **57**: 32–46.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S0305750X13002404*

Hidrobo, M. and Fernald, L. (2013). Cash Transfers and Domestic Violence, *Journal of Health Economics* **32**(1): 304–319.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S0167629612001750*

Ichino, A., Lindström, E.-A. and Viviano, E. (2014). Hidden Consequences of a First-Born Boy for Mothers, *Economics Letters* **123**(3): 274–278.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S0165176514000913*

Imai, K. and Ratkovic, M. (2013). Estimating Treatment Effect Heterogeneity in Randomized Program Evaluation, *Annals of Applied Statistics* **7**(1): 443–470.
URL: *http://projecteuclid.org/euclid.aoas/1365527206*

Iregui-Bohórquez, A. M., Ramírez-Giraldo, M. T. and Tribín-Uribe, A. M. (2019). Domestic Violence Against Rural Women in Colombia: The Role of Labor Income, *Feminist Economics* **25**(2): 146–172.
URL: *https://doi.org/10.1080/13545701.2019.1566752*

Kaushal, N. and Muchomba, F. M. (2018). Missing Time with Parents: Son Preference among Asians in the USA, *Journal of Population Economics* **31**(2): 397–427.
URL: *http://link.springer.com/10.1007/s00148-017-0668-6*

Kishor, S. and Kiersten, J. (2004). Profiling Domestic Violence - A Multi-Country Study.

Lee, S. and Shaikh, A. M. (2014). Multiple Testing and Heterogeneous Treatment Effects: Re-evaluating the Effect of Progresa on School Enrollment, *Journal of Applied Econometrics* **29**(4): 612–626.
URL: *http://doi.wiley.com/10.1002/jae.2327*

Lundberg, S., McLanahan, S. and Rose, E. (2007). Child Gender and Father Involvement in Fragile Families, *Demography* **44**(1): 79–92.
URL: *http://link.springer.com/10.1353/dem.2007.0007*

Mammen, K. (2011). Fathers' Time Investments in Children: Do Sons Get More?, *Journal of Population Economics* **24**(3): 839–871.
URL: *http://link.springer.com/10.1007/s00148-009-0272-5*

Ministerio de Salud y Protección Social y Profamilia [Producers] (2005). Colombia Salud Sexual y Reproductiva: Resultados Encuesta Nacional de Demografía y Salud 2005. [Dataset].COIR53DT.DTA., *Bogotá, Colombia: Profamilia* (ICF [Distribution], 2005).

Ministerio de Salud y Protección Social y Profamilia [Producers] (2011). Colombia Encuesta Nacional de Demografía y Salud 2010. [Dataset].COIR61DT.DTA., *Bogotá, Colombia: Profamilia* (ICF [Distribution], 2011).

Ministerio de Salud y Protección Social y Profamilia [Producers] (2017). Colombia Encuesta Nacional de Demográfia y Salud 2015. [Dataset].COIR72DT.DTA., *Bogotá, Colombia: Profamilia* (ICF [Distribution], 2017).

Nie, X. and Wager, S. (2017). Quasi-Oracle Estimation of Heterogeneous Treatment Effects.
URL: *http://arxiv.org/abs/1712.04912*

Novella, R. (2019). Parental Education, Gender Preferences and Child Nutritional Status in Peru, *Oxford Development Studies* **47**(1): 29–47.
URL: *https://www.tandfonline.com/doi/full/10.1080/13600818.2018.1495703*

Oecd (2016). Social Institutions and Gender, *Technical report*.

Oestby, G., Leiby, M. and Nordås, R. (2019). The Legacy of Wartime Violence on Intimate-Partner Abuse: Microlevel Evidence from Peru, 1980 - 2009, *International Studies Quarterly* **63**(1): 1–14.
URL: *https://academic.oup.com/isq/article/63/1/1/5290058*

Powers, S., Qian, J., Jung, K., Schuler, A., Shah, N. H., Hastie, T. and Tibshirani, R. (2018). Some Methods for Heterogeneous Treatment Effect Estimation in High Dimensions, *Statistics in Medicine* **37**(11): 1767–1787.
URL: *http://doi.wiley.com/10.1002/sim.7623*

Raley, S. and Bianchi, S. (2006). Sons, Daughters, and Family Processes: Does Gender of Children Matter?, *Annual Review of Sociology* **32**(1): 401–421.
URL: *http://www.annualreviews.org/doi/10.1146/annurev.soc.32.061604.123106*

Ribero, R. and Sánchez, F. (2005). Determinants, Effects and Costs of Domestic Violence, *Documento, CEDE* **38**.

Rosenbaum, P. R. and Rubin, D. B. (2006). The Central Role of the Propensity Score in Observational Studies for Causal Effects, *Matched Sampling for Causal Effects* **70**(1): 170–184.
URL: *https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/70.1.41*

Rubin, D. B. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies, *Journal of Educational Psychology* **66**(5): 688–701.
URL: *http://content.apa.org/journals/edu/66/5/688*

Runfola, D; Seitz, L; Hobbs, L; Panginaban, L; Oberman, R. (2020). geoBoundaries Global Administrative Database.
URL: *http://www.geoboundaries.org*

Semahegn, A. and Mengistie, B. (2015). Domestic Violence Against Women and Associated Factors in Ethiopia; Systematic Review, *Reproductive Health* **12**(1): 78.
URL: *http://reproductive-health-journal.biomedcentral.com/articles/10.1186/s12978-015-0072-1*

Sexual Health Team, D. o. H. (2015). Assessment of Termination of Pregnancy on Grounds of the Sex of the Foetus - Response to Serious Crime Act 2015, *Technical report*.

Shafer, E. F. and Malhotra, N. (2011). The Effect of a Child's Sex on Support for Traditional Gender Roles, *Social Forces* **90**(1): 209–222.
URL: *https://academic.oup.com/sf/article-lookup/doi/10.1093/sf/90.1.209*

Sivak, E. and Smirnov, I. (2019). Parents Mention Sons More Often Than Daughters on Social Media, *Proceedings of the National Academy of Sciences of the United States of America* **116**(6): 2039–2041.
URL: *http://www.pnas.org/lookup/doi/10.1073/pnas.1804996116*

Somville, V. (2020). Having a Daughter Reduces Male Violence Against a Partner, *SSRN Electronic Journal* .

Tracey, T. J. G. and Hopkins, N. (2001). Correspondence of Interests and Abilities with Occupational Choice, *Journal of Counseling Psychology* **48**(2): 178–189.
URL: *http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-0167.48.2.178*

Trivers, R. L. and Willard, D. E. (1973). Natural Selection of Parental Ability to Vary the Sex Ratio of Offspring, *Science* **179**(4068): 90–92.
URL: *http://www.sciencemag.org/cgi/doi/10.1126/science.179.4068.90*

United Nations (2020). Policy Brief: The Impact of COVID-19 on Women, *Technical report*.

UNODC Vienna (2019). Global Study on Homicide 2019, *Technical report*.

Vyas, S. and Watts, C. (2009). How Does Economic Empowerment Affect Women's Risk of Intimate Partner Violence in Low and Middle Income Countries? A Systematic Review of Published Evidence, *Journal of International Development* **21**(5): 577–602.
URL: *http://doi.wiley.com/10.1002/jid.1500*

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.
URL: *https://www.tandfonline.com/doi/full/10.1080/01621459.2017.1319839*

Wang, V., Hatlebakk, M., Tønnessen, L., Mæstad, O. and Telle, K. (2019). Gendercide and Marginalisation: An Initial Review of the Knowledge Base, *CMI Report* .

Willke, R. J., Zheng, Z., Subedi, P., Althin, R. and Mullins, C. D. (2012). From Concepts, Theory, and Evidence of Heterogeneity of Treatment Effects to Methodological Approaches: A primer, *BMC Medical Research Methodology* **12**(1): 185.
URL: *https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-12-185*

List of Figures

List of Tables

Appendix

Additional Summary Statistics

Sample Composition	Reduction of Sample	
	Absolute	Step by Step
has no living son	68395	68395
has never been in a union/marriage/relationship	46793	4236
respondent has twins	7	7
missing outcomes	48343	1153
missing main covariates	72253	12870

Table 19: Sample composition

The table shows the sample composition and how each variable impacts the number of observations in the main sample. The original data set contains N = 133,583 observations and the main sample N = 46,922. The first column shows the absolute counts for each of the variables in the original data set that do not match the condition I imposed. Starting from the top, the second column shows the number of observations that are dropped for each additional constraint.

	q5	mean	median	q95	missings	N
Characteristics of the household						
number of children 0-5 years	0.00	0.70	1.00	2.00	0.00	46922
number of children 6-14 years	0.00	1.05	1.00	3.00	0.00	46922
number of women 13-49 years	1.00	1.50	1.00	3.00	0.00	46922
number of members	3.00	4.97	5.00	9.00	0.00	46922
household head is female	0.00	0.18	0.00	1.00	0.00	46922
Characteristics of the female respondent						
is divorced	0.00	0.00	0.00	0.00	0.00	46922
lives with a partner	0.00	0.58	1.00	1.00	0.00	46922
is married	0.00	0.37	0.00	1.00	0.00	46922
is separated	0.00	0.05	0.00	0.00	0.00	46922
is widowed	0.00	0.00	0.00	0.00	0.00	46922
had first child before getting married	0.00	0.07	0.00	1.00	7671.00	46922
index on autonomy in relationship	0.00	2.51	2.00	5.00	0.00	46922
is literate	0.00	0.71	1.00	1.00	28729.00	46922
number of dead children	0.00	0.09	0.00	1.00	0.00	46922
number of living children	1.00	2.71	2.00	6.00	0.00	46922
number of unions	1.00	1.24	1.00	2.00	0.00	46922
age of partner	24.00	39.37	39.00	56.00	0.00	46922
is pregnant	0.00	0.03	0.00	0.00	0.00	46922
is the household head	0.00	0.12	0.00	1.00	0.00	46922
is wife of household head	0.00	0.75	1.00	1.00	0.00	46922
ever terminated a pregnancy	0.00	0.27	0.00	1.00	0.00	46922
is currently working	0.00	0.54	1.00	1.00	0.00	46922

Table 20: Summary statistics of additional variables in main data set

The percentiles and the mean are weighted with the domestic violence sample weight provided by DHS. The sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. The autonomy index is self-imputed based on five variables that indicate how responsibilities are shared in the relationship and ranges from 0 to 5, whereby 5 indicates that the respondent has full autonomy in all five categories. The variables included in this table are not included in the main analysis due to potential endogeneity concerns.

Propensity Score

	Entire Sample		Main Sample		p-value
	Mean	N	Mean	N	
Has experienced at least one of the three outcomes					
ever	0.37	31265	0.30	46922	0.00
before last year	0.23	31265	0.23	46922	1.00
during last year	0.25	31265	0.18	46922	0.00
Outcome: Less severe violence					
ever	0.35	31265	0.28	46922	0.00
before last year	0.20	31265	0.20	46922	1.00
during last year	0.23	31265	0.17	46922	0.00
Outcome: Severe violence					
ever	0.15	31265	0.10	46922	0.00
before last year	0.10	31265	0.08	46922	0.00
during last year	0.11	31265	0.06	46922	0.00
Outcome: Sexual violence					
ever	0.11	31265	0.07	46922	0.00
before last year	0.06	31265	0.04	46922	0.00
during last year	0.08	31265	0.04	46922	0.00

Table 21: Difference in outcomes for different samples

The mean and the standard deviation are weighted with the domestic violence sample weight provided by DHS and p-values are adjusted for multiple hypothesis testing using the bonferroni method. The entire sample is restricted to women with at least one living child who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables. The main sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. Less severe violence indicates whether the woman has ever experienced at least one of the following items: pushed, shook or had something thrown; slapped; punched with fist or hit by something harmful. Severe violence indicates whether the woman has ever experienced at least one of the following items: kicked or dragged; strangled or burnt; threatened or attacked with weapon. Sexual violence indicates whether the woman was ever physically forced into unwanted sex. All outcomes only refer to actions committed by the husband or partner.

	Entire Sample		Main Sample		p-value
	Mean	N	Mean	N	
Characteristics of the household					
number of children 0-5 years	0.65	31265	0.70	46922	0.00
number of children 6-14 years	0.74	31265	1.05	46922	0.00
number of women 13-49 years	1.75	31265	1.50	46922	0.00
number of members	4.63	31265	4.97	46922	0.00
household head is female	0.46	31265	0.18	46922	0.00
lives in urban environment	0.80	31265	0.73	46922	0.00
wealth quintile of household	3.09	31265	2.92	46922	0.00
Characteristics of the female respondent					
experienced domestic violence as child	0.35	31265	0.35	46922	0.33
age difference to partner	4.31	16425	4.28	46922	1.00
age	34.48	31265	35.10	46922	0.00
age at first child	21.04	31265	20.35	46922	0.00
was minor at first sexual intercourse	0.69	28665	0.71	46922	0.00
is divorced	0.01	31265	0.00	46922	0.00
lives with a partner	0.31	31265	0.58	46922	0.00
is married	0.19	31265	0.37	46922	0.00
is separated	0.42	31265	0.05	46922	0.00
is widowed	0.06	31265	0.00	46922	0.00
had first child before getting married	0.07	26918	0.07	39251	1.00
index on autonomy in relationship	2.96	31265	2.51	46922	0.00
is literate	0.72	9733	0.71	18193	0.64
number of dead children	0.08	31265	0.09	46922	0.00
number of living children	2.09	31265	2.71	46922	0.00
number of unions	1.21	31265	1.24	46922	0.00
works in agriculture	0.03	31265	0.04	46922	0.00
works in clerus	0.08	31265	0.07	46922	0.00
works in sales	0.24	31265	0.24	46922	1.00
works in services	0.40	31265	0.38	46922	0.00
works in skilled & manual job	0.05	31265	0.05	46922	0.00
works in technical & managerial job	0.09	31265	0.08	46922	0.00
works in unskilled & manual job	0.02	31265	0.02	46922	0.76
age of partner	36.95	16425	39.37	46922	0.00
is pregnant	0.03	31265	0.03	46922	1.00
is the household head	0.33	31265	0.12	46922	0.00
is wife of household head	0.39	31265	0.75	46922	0.00
ever terminated a pregnancy	0.25	31265	0.27	46922	0.00
total years of education	8.88	31265	8.27	46922	0.00
is currently working	0.64	31265	0.54	46922	0.00

Table 22: Difference in covariates for different samples

The mean and the standard deviation are weighted with the domestic violence sample weight provided by DHS and p-values are adjusted for multiple hypothesis testing using the bonferroni method. The entire sample is restricted to women with at least one living child who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables. The main sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. The autonomy index is self-imputed based on five variables that indicate how responsibilities are shared in the relationship and ranges from 0 to 5, whereby 5 indicates that the respondent has full autonomy in all five categories. The variable "experienced domestic violence as child" refers to actions committed by the father against the mother during childhood. "Minor" refers to all women younger than 18 years.

Variable	Relative Importance
age	0.53
total years of education	0.25
age at first child	0.15
wealth quintile of household	0.02
age difference to partner	0.01
lives in urban environment	0.01
works in technical & managerial job	0.00
works in services	0.00
was minor at first sexual intercourse	0.00
works in sales	0.00
works in clerus	0.00
works in agriculture	0.00
experienced domestic violence as child	0.00
works in skilled & manual job	0.00
works in unskilled & manual job	0.00

Table 23: Propensity score: Sorted measure of variable importance

The propensity score is estimated with a regression forest using tuned parameters (min.node.size is fixed to 1000) and includes fixed effects on a year and department level, clusters as adviced by DHS and the domestic violence sample weight provided by DHS. The variable importance indicates how often a variable was used as splitting variable in the regression forest. The continuous variables are scaled. The sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. The variable "experienced domestic violence as child" refers to actions committed by the father against the mother during childhood. "Minor" refers to all women younger than 18 years.

The results presented in Table 23 are very similar to the ones in Table 5 in Section 4.3, even though the measure of variable importance has to be interpreted with caution. There is a set of three variables - age, total years of education and age at first child - that seems to be more related to the probability of being treated than the remaining variables, which questions the assumption of exogeneity of the treatment based on observables.

Additional Explanations to the Causal Forest and Regression Forest Algorithm

Average Treatment Effect

The sample average treatment effect is the difference of the average outcome for treated and untreated observations within each subgroup:

$$\hat{\tau} = \frac{1}{|I_{1,q}|} \sum_{i \epsilon I_{1,q}} Y_i - \frac{1}{|I_{0,q}|} \sum_{i \epsilon I_{0,q}} Y_i \tag{14}$$

with $I_{w,q} := \{i | i \epsilon W_i, i \epsilon I_q\}$ and I_q is the subset of individuals whose predicted treatment effect is in the q^{th} n-tile.

The package grf allows to estimate the augmented inverse-propensity weighted average treatment effect (AIPW ATE) based on Equation 15 and its standard error (Athey and Wager, 2019).

$$\hat{\tau} = \frac{1}{|I_q|} \sum_{i \epsilon I_q} \hat{\tau}^{-i}(X_i) + \frac{W_i - \hat{e}^{-i}(X_i)}{\hat{e}^{-i}(X_i)(1 - \hat{e}^{-i}(X_i))}(Y_i - \hat{\mu}_{w_i}(X_i)) \tag{15}$$

$\hat{e}^{-i}(X_i)$ and $\hat{\tau}^{-i}(X_i)$ are based on Equations 1 and 13 and are out-of-bag estimates, meaning that the observation i was not used to estimate the functions. $\hat{\mu}_{w_i}(X_i)$ is based on Equation 6. The AIPW ATE estimator is doubly robust. Inverse propensity score weighting assigns observations that are less likely to appear in the group they are in a higher weight, which is determined by the value of their propensity score, which itself is based on observable characteristics. In general, when using inverse propensity score weighting, the weight is $\frac{1}{\hat{e}(X_i)}$ for treated individuals and $\frac{1}{1-\hat{e}(X_i)}$ for untreated individuals. If, for a certain set of characteristics, the probability to be treated is very high, individuals who have the set of characteristics, but are not treated, receive a relatively higher weight compared to individuals who have the same set of characteristics, but are treated.

The "augmented" in AIPW ATE refers to the term $(Y_i - \hat{\mu}_{w_i}(X_i))$, which is larger the larger the extent to which the estimated treatment effect for an individual differs from the average treatment effect, which depends on the difference of the predicted outcome, estimated based on the observables, to the actual outcome. If $\hat{\mu}_{w_i}(X_i)$ is specified correctly with respect to X_i, the difference $(Y_i - \hat{\mu}_{w_i}(X_i))$ shows a different treatment effect for the individual and therefore, the estimated treatment effect for the individual is larger than average in absolute size by the factor $\frac{W_i - \hat{e}^{-i}(X_i)}{\hat{e}^{-i}(X_i)(1-\hat{e}^{-i}(X_i))}$.

Tuning

The grf package chooses tuning parameter by cross-validating on the R-loss based on Equation 16 (Nie and Wager, 2017). Cross-validation is applied in the following way: Several causal forests are trained with different values for the tuning parameters and the final values are those that result in the out-of-bag estimates of the objective that minimize the R-loss. Here, $\tau()$ refers to the heterogeneous treatment effect function, $e(x)$ and $m(x)$ are estimated based on Equations 1 and 5, respectively.

$$\hat{L}_n(\tau) = \frac{1}{n}\sum_{i=1}^{n}\left(\left(Y_i - \hat{m}^{(-i)}(X_i)\right) - \left(W_i - \hat{e}^{(-i)}(X_i)\right)\tau(X_i)\right)^2 \tag{16}$$

The parameters that can be tuned in the causal forest algorithm are the following: sample.fraction (fraction of the data used to build each tree), mtry (number of variables tried for each split), min.node.size (minimum number of observations in each tree leaf), honesty.fraction (fraction of the data to determine splitting rules), honesty.prune.leaves (if estimation sample trees are subject to pruning such that no final leaf is empty), alpha (controls maximum imbalance of a split), imbalance.penalty (degree of penalty of imbalanced splits)

Regression forest

The regression forest in the grf package relies on similar principles as the causal forest, which are explained in Section 5.2 and can also be tuned, but the overall goal is conceptually different. Instead of aiming at estimating the treatment effect, the regression forest can be used to predict an outcome based on the conditional mean function in Equation 6.

$$\hat{\mu}(x) = \sum_{i=1}^{n} Y_i \alpha_i(x) \tag{17}$$

Here, $\alpha_i(x) = \frac{1}{B}\sum_{b=1}^{B}\frac{(\{X_i \epsilon L_b(x)\})}{|\{i : X_i \epsilon L_b(x)\}|}$ with B as the total number of trees in the forest and $L_b(x)$ as the leaf where x falls into in tree b. The predicted outcome $\hat{\mu}(x)$ is the mean of the outcomes of all observations in the training sample, weighted by the adaptive weights $\alpha_i(x)$ learned through the regression forest. The weight is higher, the closer an observation i is to the observation with $X_i = x$, defined over how often the observations fall in the same final leaf across all B trees. The trees used to build the regression forest are honest regression trees.

Additional Results of LPM, Logit and Probit

	Linear probability model			Logit			Probit		
Variables	est	s.e	p-value	est	s.e	p-value	est	s.e	p-value
lives in urban environment	0.02	0.01	0.07	0.13	0.04	0.02	0.13	0.02	0.00
wealth quintile of household	-0.00	0.00	1.00	-0.01	0.01	1.00	-0.01	0.01	1.00
experienced domestic violence as child	0.08	0.01	0.00	0.44	0.02	0.00	0.44	0.01	0.00
age	0.06	0.00	0.00	0.37	0.01	0.00	0.37	0.01	0.00
age at first child	-0.03	0.01	0.00	-0.21	0.02	0.00	-0.21	0.01	0.00
works in agriculture	0.05	0.01	0.00	0.28	0.06	0.00	0.28	0.03	0.00
works in sales	0.04	0.01	0.00	0.24	0.04	0.00	0.24	0.02	0.00
works in services	0.04	0.01	0.00	0.25	0.03	0.00	0.25	0.02	0.00
works in skilled & manual job	0.05	0.01	0.03	0.40	0.07	0.00	0.40	0.04	0.00
works in technical & managerial job	0.00	0.01	1.00	0.03	0.06	1.00	0.03	0.03	1.00
works in unskilled & manual job	0.03	0.02	1.00	0.21	0.09	1.00	0.21	0.05	0.00
total years of education	0.00	0.00	1.00	0.06	0.02	0.04	0.06	0.01	0.00
age difference to partner	0.00	0.00	0.64	0.04	0.01	0.02	0.04	0.01	0.00
was minor at first sexual intercourse	0.02	0.00	0.00	0.07	0.02	0.00	0.07	0.01	0.00
treatment	-0.01	0.01	1.00	-0.08	0.03	0.05	-0.08	0.01	0.00

Table 24: Regression results: Less severe violence before last year

The sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. The continuous variables are scaled and fixed effects are on a year and department level. Standard errors are robust and clustered as adviced by DHS and p-values are adjusted for multiple hypothesis testing based on the bonferroni method. The observations are weighted with the domestic violence sample weight provided by DHS. Less severe violence indicates whether the woman has experienced at least one of the following items before the last year: pushed, shook or had something thrown; slapped; punched with fist or hit by something harmful. The outcome only refers to actions committed by the husband or partner. The variable "experienced domestic violence as child" refers to actions committed by the father against the mother during childhood. "Minor" refers to all women younger than 18 years.

	Linear probability model			Logit			Probit		
Variables	est	s.e	p-value	est	s.e	p-value	est	s.e	p-value
lives in urban environment	0.02	0.01	0.00	0.23	0.05	0.00	0.23	0.03	0.00
wealth quintile of household	-0.01	0.00	0.09	-0.04	0.02	1.00	-0.04	0.01	0.00
experienced domestic violence as child	0.05	0.00	0.00	0.57	0.04	0.00	0.57	0.02	0.00
age	0.03	0.00	0.00	0.42	0.02	0.00	0.42	0.01	0.00
age at first child	-0.01	0.01	0.76	-0.26	0.03	0.00	-0.26	0.01	0.00
works in agriculture	0.02	0.01	1.00	0.25	0.09	0.19	0.25	0.04	0.00
works in sales	0.03	0.01	0.00	0.31	0.06	0.00	0.31	0.03	0.00
works in services	0.03	0.00	0.00	0.38	0.05	0.00	0.38	0.03	0.00
works in skilled & manual job	0.03	0.01	0.16	0.53	0.10	0.00	0.53	0.05	0.00
works in technical & managerial job	0.00	0.01	1.00	0.06	0.10	1.00	0.06	0.05	1.00
works in unskilled & manual job	0.02	0.01	1.00	0.35	0.13	0.34	0.35	0.07	0.00
total years of education	-0.01	0.00	0.03	-0.08	0.02	0.08	-0.08	0.01	0.00
age difference to partner	0.00	0.00	0.21	0.05	0.02	0.13	0.05	0.01	0.00
was minor at first sexual intercourse	0.01	0.00	0.01	0.11	0.02	0.00	0.11	0.01	0.00
treatment	-0.00	0.00	1.00	-0.07	0.04	1.00	-0.07	0.02	0.00

Table 25: Regression results: Severe violence before last year

The sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. The continuous variables are scaled and fixed effects are on a year and department level. Standard errors are robust and clustered as adviced by DHS and p-values are adjusted for multiple hypothesis testing based on the bonferroni method. The observations are weighted with the domestic violence sample weight provided by DHS. Severe violence indicates whether the woman has experienced at least one of the following items before the last year: kicked or dragged; strangled or burnt; threatened or attacked with weapon. The outcome only refers to actions committed by the husband or partner. The variable "experienced domestic violence as child" refers to actions committed by the father against the mother during childhood. "Minor" refers to all women younger than 18 years.

	Linear probability model			Logit			Probit		
Variables	est	s.e	p-value	est	s.e	p-value	est	s.e	p-value
lives in urban environment	0.01	0.00	0.02	0.32	0.07	0.00	0.32	0.03	0.00
wealth quintile of household	-0.00	0.00	0.58	-0.07	0.03	1.00	-0.07	0.01	0.00
experienced domestic violence as child	0.03	0.00	0.00	0.65	0.05	0.00	0.65	0.02	0.00
age	0.02	0.00	0.00	0.54	0.03	0.00	0.54	0.01	0.00
age at first child	-0.01	0.00	0.00	-0.22	0.04	0.00	-0.22	0.02	0.00
works in agriculture	0.02	0.01	0.06	0.44	0.13	0.03	0.44	0.06	0.00
works in sales	0.02	0.00	0.00	0.33	0.09	0.01	0.33	0.04	0.00
works in services	0.01	0.00	0.00	0.36	0.08	0.00	0.36	0.03	0.00
works in skilled & manual job	0.02	0.01	0.75	0.52	0.14	0.01	0.52	0.06	0.00
works in technical & managerial job	-0.00	0.00	1.00	0.11	0.14	1.00	0.11	0.06	1.00
works in unskilled & manual job	-0.00	0.01	1.00	0.13	0.20	1.00	0.13	0.09	1.00
total years of education	-0.00	0.00	1.00	0.01	0.03	1.00	0.01	0.02	1.00
age difference to partner	0.00	0.00	1.00	0.06	0.02	0.71	0.06	0.01	0.00
was minor at first sexual intercourse	0.00	0.00	1.00	0.06	0.03	1.00	0.06	0.01	0.00
treatment	-0.00	0.00	1.00	-0.09	0.05	1.00	-0.09	0.02	0.01

Table 26: Regression results: Sexual violence before last year

The sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. The continuous variables are scaled and fixed effects are on a year and department level. Standard errors are robust and clustered as adviced by DHS and p-values are adjusted for multiple hypothesis testing based on the bonferroni method. The observations are weighted with the domestic violence sample weight provided by DHS. Sexual violence indicates whether the woman was physically forced into unwanted sex before the last year. The outcome only refers to actions committed by the husband or partner. The variable "experienced domestic violence as child" refers to actions committed by the father against the mother during childhood. "Minor" refers to all women younger than 18 years.

	Linear probability model			Logit			Probit		
Variables	est	s.e	p-value	est	s.e	p-value	est	s.e	p-value
lives in urban environment	0.06	0.01	0.00	0.36	0.04	0.00	0.36	0.02	0.00
wealth quintile of household	-0.02	0.00	0.00	-0.13	0.02	0.00	-0.13	0.01	0.00
experienced domestic violence as child	0.07	0.01	0.00	0.51	0.03	0.00	0.51	0.01	0.00
age	-0.04	0.00	0.00	-0.31	0.01	0.00	-0.31	0.01	0.00
age at first child	-0.02	0.00	0.00	-0.13	0.02	0.00	-0.13	0.01	0.00
works in agriculture	0.02	0.01	1.00	0.22	0.07	0.04	0.22	0.04	0.00
works in sales	0.04	0.01	0.00	0.31	0.04	0.00	0.31	0.02	0.00
works in services	0.04	0.01	0.00	0.35	0.04	0.00	0.35	0.02	0.00
works in skilled & manual job	0.05	0.01	0.01	0.44	0.07	0.00	0.44	0.04	0.00
works in technical & managerial job	0.01	0.01	1.00	0.11	0.07	1.00	0.11	0.04	0.19
works in unskilled & manual job	0.04	0.02	0.70	0.28	0.10	0.16	0.28	0.05	0.00
total years of education	-0.01	0.00	0.35	-0.06	0.02	0.05	-0.06	0.01	0.00
age difference to partner	-0.02	0.00	0.00	-0.16	0.01	0.00	-0.16	0.01	0.00
was minor at first sexual intercourse	0.01	0.00	0.00	0.10	0.02	0.00	0.10	0.01	0.00
treatment	-0.01	0.00	1.00	-0.08	0.03	0.20	-0.08	0.02	0.00

Table 27: Regression results: Less severe violence last year

The sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. The continuous variables are scaled and fixed effects are on a year and department level. Standard errors are robust and clustered as adviced by DHS and p-values are adjusted for multiple hypothesis testing based on the bonferroni method. The observations are weighted with the domestic violence sample weight provided by DHS. Less severe violence indicates whether the woman has experienced at least one of the following items during the last year: pushed, shook or had something thrown; slapped; punched with fist or hit by something harmful. The outcome only refers to actions committed by the husband or partner. The variable "experienced domestic violence as child" refers to actions committed by the father against the mother during childhood. "Minor" refers to all women younger than 18 years.

	Linear probability model			Logit			Probit		
Variables	est	s.e	p-value	est	s.e	p-value	est	s.e	p-value
lives in urban environment	0.03	0.00	0.00	0.38	0.06	0.00	0.38	0.03	0.00
wealth quintile of household	-0.01	0.00	0.00	-0.21	0.03	0.00	-0.21	0.01	0.00
experienced domestic violence as child	0.03	0.00	0.00	0.49	0.04	0.00	0.49	0.02	0.00
age	-0.01	0.00	0.00	-0.17	0.02	0.00	-0.17	0.01	0.00
age at first child	-0.01	0.00	0.00	-0.22	0.03	0.00	-0.22	0.01	0.00
works in agriculture	0.01	0.01	1.00	0.30	0.10	0.11	0.30	0.05	0.00
works in sales	0.02	0.00	0.00	0.36	0.07	0.00	0.36	0.03	0.00
works in services	0.02	0.00	0.00	0.44	0.06	0.00	0.44	0.03	0.00
works in skilled & manual job	0.04	0.01	0.05	0.61	0.11	0.00	0.61	0.05	0.00
works in technical & managerial job	0.01	0.00	0.31	0.13	0.12	1.00	0.13	0.05	1.00
works in unskilled & manual job	0.03	0.01	0.29	0.41	0.14	0.26	0.41	0.07	0.00
total years of education	-0.01	0.00	0.01	-0.14	0.03	0.00	-0.14	0.01	0.00
age difference to partner	-0.01	0.00	0.01	-0.12	0.02	0.00	-0.12	0.01	0.00
was minor at first sexual intercourse	0.00	0.00	0.73	0.13	0.03	0.00	0.13	0.01	0.00
treatment	-0.00	0.00	1.00	-0.09	0.04	1.00	-0.09	0.02	0.00

Table 28: Regression results: Severe violence last year

The sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. The continuous variables are scaled and fixed effects are on a year and department level. Standard errors are robust and clustered as adviced by DHS and p-values are adjusted for multiple hypothesis testing based on the bonferroni method. The observations are weighted with the domestic violence sample weight provided by DHS. Severe violence indicates whether the woman has experienced at least one of the following items during the last year: kicked or dragged; strangled or burnt; threatened or attacked with weapon. The outcome only refers to actions committed by the husband or partner. The variable "experienced domestic violence as child" refers to actions committed by the father against the mother during childhood. "Minor" refers to all women younger than 18 years.

	Linear probability model			Logit			Probit		
Variables	est	s.e	p-value	est	s.e	p-value	est	s.e	p-value
lives in urban environment	0.01	0.00	0.11	0.22	0.07	0.04	0.22	0.03	0.00
wealth quintile of household	-0.01	0.00	0.00	-0.17	0.03	0.00	-0.17	0.01	0.00
experienced domestic violence as child	0.02	0.00	0.00	0.46	0.05	0.00	0.46	0.02	0.00
age	0.00	0.00	0.69	0.08	0.02	0.07	0.08	0.01	0.00
age at first child	-0.01	0.00	0.00	-0.21	0.03	0.00	-0.21	0.01	0.00
works in agriculture	0.02	0.01	0.46	0.33	0.11	0.18	0.33	0.05	0.00
works in sales	0.01	0.00	0.19	0.24	0.08	0.07	0.24	0.03	0.00
works in services	0.01	0.00	0.00	0.37	0.07	0.00	0.37	0.03	0.00
works in skilled & manual job	0.02	0.01	0.19	0.46	0.13	0.02	0.46	0.06	0.00
works in technical & managerial job	0.01	0.00	0.66	0.13	0.13	1.00	0.13	0.06	1.00
works in unskilled & manual job	0.01	0.01	1.00	0.36	0.16	1.00	0.36	0.07	0.00
total years of education	-0.00	0.00	0.11	-0.05	0.03	1.00	-0.05	0.02	0.08
age difference to partner	-0.00	0.00	1.00	-0.02	0.02	1.00	-0.02	0.01	1.00
was minor at first sexual intercourse	0.00	0.00	0.22	0.09	0.03	0.30	0.09	0.01	0.00
treatment	0.00	0.00	1.00	-0.00	0.05	1.00	-0.00	0.02	1.00

Table 29: Regression results: Sexual violence last year

The sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. The continuous variables are scaled and fixed effects are on a year and department level. Standard errors are robust and clustered as adviced by DHS and p-values are adjusted for multiple hypothesis testing based on the bonferroni method. The observations are weighted with the domestic violence sample weight provided by DHS. Sexual violence indicates whether the woman was physically forced into unwanted sex during the last year. The outcome only refers to actions committed by the husband or partner. The variable "experienced domestic violence as child" refers to actions committed by the father against the mother during childhood. "Minor" refers to all women younger than 18 years.

	Sample ATE	AIPW ATE
Quartile 1	-0.012821	-0.042678
	(0.008113)	(0.010719)
Quartile 2	-0.004239	0.013742
	(0.008123)	(0.009142)
Quartile 3	-0.004081	0.005407
	(0.008191)	(0.008438)
Quartile 4	-0.005787	-0.013106
	(0.008349)	(0.008648)
p-value	0.8554	3e-04

Table 30: Additional outcome: ATE estimates within quartiles of treatment effect

The quartiles on the y-axis are the quartiles of the predicted treatment effect, the ATE per subgroup is defined by the out-of-bag CATE. The sample ATE is the difference of the average outcome between control and treatment group in the treatment effect quartile. The AIPW ATE is the augmented inverse-propensity weighted average treatment effect in the treatment effect quintile. The p-values are testing the null hypothesis: ATE is constant across the quartiles. The sample ATE uses an F-Test and the AIPW ATE uses a Wald test. The causal forest is estimated using 0.8 of the sample as training sample with input parameters tuned once on the training sample and the domestic violence sample weight provided by DHS. The number of trees equals the number of women in the training sample (N = 37,537). The sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. Less severe violence indicates whether the woman has experienced at least one of the following items during the last year: pushed, shook or had something thrown; slapped; punched with fist or hit by something harmful. The outcome only refers to actions committed by the husband or partner during the last year.

Additional Results for Heterogeneous Treatment Effects

In the following, I will present results of the causal forest for the outcome whether a woman experienced less severe domestic violence during the last year.

In Figure 13, the range of values is smaller compared to Figure 6, which hints towards a weaker heterogeneity in treatment effects for the same outcome during the last year compared to ever. The values of the distribution are almost entirely negative, which can indicate that the subgroup for which the treatment effect might have been positive is smaller compared to the general outcome presented in Section 6.2.

The average treatment effects across all quartiles in Table 30 are almost entirely negative, regardless of the estimation method. The AIPW ATE is statistically significantly different from zero on the 1% significance level in the first quartile of the predicted treatment effect. The difference in the average treatment effects across the quartiles is statistically significantly different on the 1% significance level for the AIPW ATE. The AIPW ATE differs between -4.3 and 1.4 percentage points over the quartiles.

Based on Table 31, the treatment effect for the women in the first quartile is different to the one for the remaining women, since the difference in the average treatment effect is statistically significantly different from the first to the third and fourth quartile on the 1% significance level. The difference in the treatment effect lies between 4.8 and 5.6 percentage points, which is similar compared to the results presented in Table 11.

Table 32 shows the difference in the covariate means across the four subgroups, based on the quartiles of the predicted treatment effect. It is important to take into account that the women in the first quartile seem to have different and stronger treatment effects compared to the women in the remaining three quartiles. However, the actual variation of the covariate values does not have to be related to heterogeneity. Compared to the results presented in Table 12, the differences between women in the fourth quartile and the other quartiles are only slightly

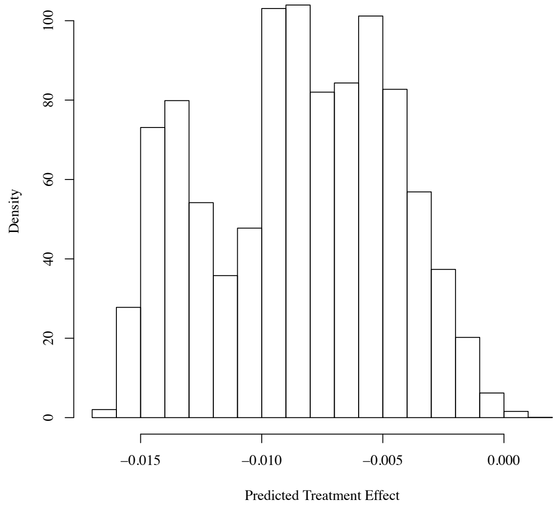

Figure 13: Additional outcome: Distribution of out-of-bag CATE

The x-axis shows the distribution of predicted treatment effects for the training sample (N = 37,537) and the y-axis the probability density. The causal forest is estimated using 0.8 of the sample as training sample with input parameters tuned once on the training sample and the domestic violence sample weight provided by DHS. The number of trees equals the number of women in the training sample. The sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. Less severe violence indicates whether the woman has experienced at least one of the following items during the last year: pushed, shook or had something thrown; slapped; punched with fist or hit by something harmful. The outcome only refers to actions committed by the husband or partner during the last year.

	Quartile 1	Quartile 2	Quartile 3	Quartile 4
Quartile 1				
Quartile 2	0.056 (0.015)			
Quartile 3	0.048 (0.015)	-0.008 (0.013)		
Quartile 4	 (0.015)	-0.027 (0.013)	-0.019 (0.012)	

Table 31: Additional outcome: Pairwise comparison of AIPW ATE estimates within quartiles of the treatment effect

The quartiles are based on the predicted treatment effect. The AIPW ATE is the augmented inverse-propensity weighted average treatment effect in the treatment effect quartile. The difference between two quartiles is the difference in the AIPW ATE, standard errors are in parentheses. The gray background color indicates that the treatment effects are statistically significantly different from each other on the 5 percent significance level, the black background color indicates the same on the 1 percent significance level. The causal forest is estimated using 0.8 of the sample as training sample with input parameters tuned once on the training sample and the domestic violence sample weight provided by DHS. The number of trees equals the number of women in the training sample (N = 37,537). The sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. Less severe violence indicates whether the woman has experienced at least one of the following items during the last year: pushed, shook or had something thrown; slapped; punched with fist or hit by something harmful. The outcome only refers to actions committed by the husband or partner during the last year.

different. The evidence for the difference in occupational fields is more extreme in terms of the distribution across the subgroups and some results do not hold anymore. For the characteristics, such as age, the years spent in education, the age difference and the wealth, the results presented in Table 32 support the findings in Table 12. Overall, the subgroups defined over the covariate values are similar for both outcomes.

Covariates	Quartile 1	Quartile 2	Quartile 3	Quartile 4
lives in urban environment	0.744	0.605	0.601	0.811
	(0.005)	(0.005)	(0.005)	(0.005)
wealth quintile of household	2.393	2.291	2.43	2.98
	(0.013)	(0.013)	(0.013)	(0.013)
experienced domestic violence as child	0.385	0.351	0.34	0.302
	(0.005)	(0.005)	(0.005)	(0.005)
age	32.69	33.89	35.66	37.13
	(0.085)	(0.085)	(0.085)	(0.085)
age at first child	18.28	19.43	20	22.05
	(0.041)	(0.041)	(0.041)	(0.041)
works in agriculture	0	0.086	0.109	0.029
	(0.002)	(0.002)	(0.002)	(0.002)
works in sales	0	0.004	0.31	0.638
	(0.003)	(0.003)	(0.003)	(0.003)
works in services	1	0.448	0.078	0.001
	(0.003)	(0.003)	(0.003)	(0.003)
works in skilled & manual job	0	0.058	0.052	0.022
	(0.002)	(0.002)	(0.002)	(0.002)
works in technical & managerial job	0	0.006	0.117	0.147
	(0.003)	(0.003)	(0.003)	(0.003)
works in unskilled & manual job	0	0.027	0.033	0.017
	(0.001)	(0.001)	(0.001)	(0.001)
total years of education	6.821	6.838	7.823	9.986
	(0.042)	(0.042)	(0.042)	(0.042)
age difference to partner	6.282	5.397	4.731	1.327
	(0.072)	(0.072)	(0.072)	(0.072)
was minor at first sexual intercourse	0.994	0.734	0.79	0.447
	(0.004)	(0.004)	(0.004)	(0.004)

Table 32: Additional outcome: Values of covariates across the quartiles of the treatment effect

The quartiles are based on the predicted treatment effect and the values indicate the average value of the covariate in the quartile. The colors indicate the position of the mean of the subgroup in the standardized empirical distribution. The standardized distribution is colored from a scale of +/- 0.9. The causal forest is estimated using 0.8 of the sample as training sample with input parameters tuned once on the training sample and the domestic violence sample weight provided by DHS. The number of trees equals the number of women in the training sample (N = 37,537). The sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. Less severe violence indicates whether the woman has experienced at least one of the following items during the last year: pushed, shook or had something thrown; slapped; punched with fist or hit by something harmful. The outcome only refers to actions committed by the husband or partner during the last year. The variable "experienced domestic violence as child" refers to actions committed by the father against the mother during childhood. "Minor" refers to all women younger than 18 years.

The four variables that were used most often for the splitting rules in the causal forest in Table 33 are the same as in Table 13. Therefore, it seems like the set of four variables, which was used in more than half of all splitting decisions is essential for the definition of the subgroups.

Variable	Relative Importance
age difference to partner	0.13
age	0.12
age at first child	0.12
total years of education	0.12
wealth quintile of household	0.09
works in services	0.08
was minor at first sexual intercourse	0.08
works in sales	0.07
lives in urban environment	0.06
experienced domestic violence as child	0.06
works in technical & managerial job	0.03
works in agriculture	0.03
works in skilled & manual job	0.02
works in unskilled & manual job	0.01

Table 33: Additional outcome: Sorted measure of variable importance

The measure of variable importance indicates how often a variable was used as splitting variable. The causal forest is estimated using 0.8 of the sample as training sample with input parameters tuned once on the training sample and the domestic violence sample weight provided by DHS. The number of trees equals the number of women in the training sample (N = 37,537). The sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. Less severe violence indicates whether the woman has experienced at least one of the following items during the last year: pushed, shook or had something thrown; slapped; punched with fist or hit by something harmful. The outcome only refers to actions committed by the husband or partner during the last year. The variable "experienced domestic violence as child" refers to actions committed by the father against the mother during childhood. "Minor" refers to all women younger than 18 years.

Additional Robustness Checks

Quintiles

Table 34 shows the sample ATE and AIPW ATE for the five subgroups based on the quintiles of the predicted treatment effect. When using the AIPW ATE, the average treatment effects are statistically significantly different from each other on the 1% significance level and range from minus four to 2.8 percentage points. The average treatment effect in the first and third quintile are statistically significantly different from zero on the 1% significance level.

	Sample ATE	AIPW ATE
Quintile 1	-0.03727	-0.039685
	(0.011078)	(0.012831)
Quintile 2	-0.020388	-0.018867
	(0.010924)	(0.012291)
Quintile 3	-0.036932	-0.031899
	(0.010975)	(0.011887)
Quintile 4	-0.023943	0.011678
	(0.010853)	(0.010781)
Quintile 5	-0.018634	0.027805
	(0.011208)	(0.011607)
p-value	0.6139	1e-04

Table 34: ATE estimates within quintiles of treatment effect

The quintiles on the y-axis are the quintiles of the treatment effect, the ATE per subgroup is defined by the out-of-bag CATE. The sample ATE is the difference of the average outcome between control and treatment group in the treatment effect quintile. The AIPW ATE is the augmented inverse-propensity weighted average treatment effect in the treatment effect quintile. The p-values are testing the null hypothesis: ATE is constant across the quintiles. The sample ATE uses an F-Test and the AIPW ATE uses a Wald test. The causal forest is estimated using 0.8 of the sample as training sample with input parameters tuned once on the training sample and the domestic violence sample weight provided by DHS. The number of trees equals the number of women in the training sample (N = 37,537). The sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. Less severe violence indicates whether the woman has ever experienced at least one of the following items: pushed, shook or had something thrown; slapped; punched with fist or hit by something harmful. The outcome only refers to actions ever committed by the husband or partner.

Table 35 shows the difference in the AIPW ATE across the quintiles. The difference in average treatment effects is statistically significantly different from zero on the 1% significance level in several cases and ranges from 4.4 percentage points to 6.7 percentage points. This indicates that the women with a strongly negative treatment effect as in the first quintile and with a positive treatment effect as in the fifth quintile do have a different treatment effect compared to the remaining women in the sample.

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Quintile 1					
Quintile 2	0.021 (0.018)				
Quintile 3	0.008 (0.018)	-0.013 (0.017)			
Quintile 4	0.051 (0.018)	(0.017)	0.044 (0.017)		
Quintile 5	0.067 (0.018)	0.047 (0.017)	0.06 (0.017)	0.016 (0.015)	

Table 35: Pairwise comparison of AIPW ATE estimates within quintiles of the treatment effect

The quintiles are based on the predicted treatment effect. The AIPW ATE is the augmented inverse-propensity weighted average treatment effect in the treatment effect quintile. The difference between two quintiles is the difference in the AIPW ATE, standard errors are in parentheses. The gray background color indicates that the treatment effects are statistically significantly different from each other on the 5 percent significance level, the black background color indicates the same on the 1 percent significance level. The causal forest is estimated using 0.8 of the sample as training sample with input parameters tuned once on the training sample and the domestic violence sample weight provided by DHS. The number of trees equals the number of women in the training sample (N = 37,537). The sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates. Less severe violence indicates whether the woman has ever experienced at least one of the following items: pushed, shook or had something thrown; slapped; punched with fist or hit by something harmful. The outcome only refers to actions ever committed by the husband or partner.

Attrition Bias

Variables	Mean	N	Mean	N	p-value
ever experienced less severe domestic violence	0.28	46928	0.47	11107	0.00
ever experienced severe domestic violence	0.10	46928	0.25	11107	0.00
ever experienced sexual domestic violence	0.07	46928	0.19	11107	0.00
experienced domestic violence as child	0.35	46928	0.34	14070	0.00
lives in urban environment	0.73	46928	0.82	14470	0.00
wealth quintile of household	2.92	46928	3.03	14470	0.00
age	35.10	46928	35.70	14470	0.00
age at first child	20.35	46928	20.54	14470	0.00
was minor at first sexual intercourse	0.71	46928	0.70	14134	0.10
has ever been in a union, marriage or relationship	1.24	46928	1.27	11107	0.00
works in agriculture	0.04	46928	0.03	14470	0.00
works in clerus	0.07	46928	0.08	14470	0.00
works in sales	0.24	46928	0.23	14470	0.03
works in services	0.38	46928	0.45	14470	0.00
works in skilled & manual job	0.05	46928	0.06	14470	0.00
works in technical & managerial job	0.08	46928	0.07	14470	0.08
works in unskilled & manual job	0.02	46928	0.02	14470	0.00
total years of education	8.27	46928	8.24	14470	1.00
treatment	0.68	46928	0.73	14470	0.00

Table 36: Attrition bias: Difference in covariates for Step 1

The mean and the standard deviation are weighted with the domestic violence sample weight provided by DHS and p-values are adjusted for multiple hypothesis testing using the bonferroni method. The entire sample is restricted to women with at least one living son. The left column includes all women that are in the main sample, which is restricted to women with at least one living son who have ever been in a relationship/union/marriage and have non-missing values for the outcome variables and the main covariates. The right column shows the women from the entire sample who have at least one living son with a missing value for the age difference in the relationship. The means in the right column are calculated excluding all missing values. Less severe violence indicates whether the woman has ever experienced at least one of the following items: pushed, shook or had something thrown; slapped; punched with fist or hit by something harmful. Severe violence indicates whether the woman has ever experienced at least one of the following items: kicked or dragged; strangled or burnt; threatened or attacked with weapon. Sexual violence indicates whether the woman was ever physically forced into unwanted sex. All outcomes only refer to actions ever committed by the husband or partner. The variable "experienced domestic violence as child" refers to actions committed by the father against the mother during childhood. "Minor" refers to all women younger than 18 years.

Variables	Mean	N	Mean	N	p-value
experienced domestic violence as child	0.35	46928		0	
lives in urban environment	0.73	46928	0.82	840	0.00
wealth quintile of household	2.92	46928	3.17	840	0.00
age	35.10	46928	30.34	840	0.00
age at first child	20.35	46928	21.14	840	0.00
was minor at first sexual intercourse	0.71	46928	0.79	840	0.00
has ever been in a union, marriage or relationship	1.24	46928		0	
works in agriculture	0.04	46928	0.04	840	1.00
works in clerus	0.07	46928	0.11	840	0.00
works in sales	0.24	46928	0.22	840	1.00
works in services	0.38	46928	0.28	840	0.00
works in skilled & manual job	0.05	46928	0.04	840	1.00
works in technical & managerial job	0.08	46928	0.18	840	0.00
works in unskilled & manual job	0.02	46928	0.06	840	0.00
total years of education	8.27	46928	10.79	840	0.00
treatment	0.68	46928	0.87	840	0.00

Table 37: Attrition bias: Difference in covariates for Step 2

The mean and the standard deviation are weighted with the domestic violence sample weight provided by DHS and p-values are adjusted for multiple hypothesis testing using the bonferroni method. The entire sample is restricted to women with at least one living son. The left column includes all women that are in the main sample, which is restricted to women with at least one living son who have ever been in a relationship/union/marriage and have non-missing values for the outcome variables and the main covariates. The right column shows the women from the entire sample who have at least one living son with a non-missing value for the age difference in the relationship, but missing values for the outcomes. The means in the right column are calculated excluding all missing values. Less severe violence indicates whether the woman has ever experienced at least one of the following items: pushed, shook or had something thrown; slapped; punched with fist or hit by something harmful. Severe violence indicates whether the woman has ever experienced at least one of the following items: kicked or dragged; strangled or burnt; threatened or attacked with weapon. Sexual violence indicates whether the woman was ever physically forced into unwanted sex. All outcomes only refer to actions ever committed by the husband or partner. The variable "experienced domestic violence as child" refers to actions committed by the father against the mother during childhood. "Minor" refers to all women younger than 18 years.

Variables	Mean	N	Mean	N	p-value
experienced domestic violence as child	0.35	46928	0.30	1764	0.00
lives in urban environment	0.73	46928	0.65	1764	0.00
wealth quintile of household	2.92	46928	2.66	1764	0.00
age	35.10	46928	38.82	1764	0.00
age at first child	20.35	46928	20.51	1764	1.00
has ever been in a union, marriage or relationship	1.24	46928	1.27	1764	0.02
works in agriculture	0.04	46928	0.10	1764	0.00
works in clerus	0.07	46928	0.04	1764	0.00
works in sales	0.24	46928	0.23	1764	1.00
works in services	0.38	46928	0.37	1764	1.00
works in skilled & manual job	0.05	46928	0.05	1764	1.00
works in technical & managerial job	0.08	46928	0.05	1764	0.00
works in unskilled & manual job	0.02	46928	0.03	1764	0.14
total years of education	8.27	46928	6.83	1764	0.00
treatment	0.68	46928	0.67	1764	1.00

Table 38: Attrition bias: Difference in covariates for Step 3

The mean and the standard deviation are weighted with the domestic violence sample weight provided by DHS and p-values are adjusted for multiple hypothesis testing using the bonferroni method. The entire sample is restricted to women with at least one living son. The left column includes all women that are in the main sample, which is restricted to women with at least one living son who have ever been in a relationship/union/marriage and have non-missing values for the outcome variables and the main covariates. The right column shows the women from the entire sample who have at least one living son with non-missing values for the age difference in the relationship and the outcomes, but a missing value for whether they were minor at their first sexual intercourse. The means in the right column are calculated excluding all missing values. Less severe violence indicates whether the woman has ever experienced at least one of the following items: pushed, shook or had something thrown; slapped; punched with fist or hit by something harmful. Severe violence indicates whether the woman has ever experienced at least one of the following items: kicked or dragged; strangled or burnt; threatened or attacked with weapon. Sexual violence indicates whether the woman was ever physically forced into unwanted sex. All outcomes only refer to actions ever committed by the husband or partner. The variable "experienced domestic violence as child" refers to actions committed by the father against the mother during childhood. "Minor" refers to all women younger than 18 years.

	Sample ATE	AIPW ATE
Quartile 1	-0.027184	-0.031603
	(0.009339)	(0.010322)
Quartile 2	-0.03164	-0.0082
	(0.009143)	(0.009742)
Quartile 3	-0.026827	-0.001583
	(0.009064)	(0.009672)
Quartile 4	-0.027215	0.001951
	(0.00908)	(0.009171)
p-value	0.9795	0.0379

Table 39: Attrition bias: ATE estimates within quartiles of treatment effect

The quartiles on the y-axis are the quartiles of the predicted treatment effect, the ATE per subgroup is defined by the out-of-bag CATE. The sample ATE is the difference of the average outcome between control and treatment group in the treatment effect quartile. The AIPW ATE is the augmented inverse-propensity weighted average treatment effect in the treatment effect quartile. The p-values are testing the null hypothesis: ATE is constant across the quartiles. The sample ATE uses an F-Test and the AIPW ATE uses a Wald test. The causal forest is estimated using 0.8 of the sample as training sample with input parameters tuned once on the training sample and the domestic violence sample weight provided by DHS. The number of trees equals the number of women in the training sample (N = 46,153). The sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates, except the age difference in relationship. In the sample, 46,922 women indicated the value of the variable. I use this sample to predict the value of the variable for the 10,770 women with a missing value using a regression forest. The regression forest takes all main covariates except the age difference in the relationship to predict the value of the age difference in the relationship. The regression forest uses the domestic violence sample weights (DHS), min.node.size is set to 1000 and all other parameters are tuned. Less severe violence indicates whether the woman has ever experienced at least one of the following items: pushed, shook or had something thrown; slapped; punched with fist or hit by something harmful. The outcome only refers to actions ever committed by the husband or partner.

	Quartile 1	Quartile 2	Quartile 3	Quartile 4
Quartile 1				
Quartile 2	0.023 (0.015)			
Quartile 3	0.03 (0.015)	0.007 (0.014)		
Quartile 4	0.034 (0.015)	0.01 (0.014)	0.004 (0.014)	

Table 40: Attrition bias: Pairwise comparison of AIPW ATE estimates within quartiles of the treatment effect

The quartiles are based on the predicted treatment effect. The AIPW ATE is the augmented inverse-propensity weighted average treatment effect in the treatment effect quartile. The difference between two quartiles is the difference in the AIPW ATE, standard errors are in parentheses. The gray background color indicates that the treatment effects are statistically significantly different from each other on the 5 percent significance level, the black background color indicates the same on the 1 percent significance level. The causal forest is estimated using 0.8 of the sample as training sample with input parameters tuned once on the training sample and the domestic violence sample weight provided by DHS. The number of trees equals the number of women in the training sample (N = 46,153). The sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates, except the age difference in relationship. In the sample, 46,922 women indicated the value of the variable. I use this sample to predict the value of the variable for the 10,770 women with a missing value using a regression forest. The regression forest takes all main covariates except the age difference in the relationship to predict the value of the age difference in the relationship. The regression forest uses the domestic violence sample weights (DHS), min.node.size is set to 1000 and all other parameters are tuned. Less severe violence indicates whether the woman has ever experienced at least one of the following items: pushed, shook or had something thrown; slapped; punched with fist or hit by something harmful. The outcome only refers to actions ever committed by the husband or partner.

Covariates	Quartile 1	Quartile 2	Quartile 3	Quartile 4
lives in urban environment	0.763	0.677	0.658	0.754
	(0.004)	(0.004)	(0.004)	(0.004)
wealth quintile of household	2.443	2.431	2.466	2.874
	(0.012)	(0.012)	(0.012)	(0.012)
experienced domestic violence as child	0.322	0.363	0.367	0.328
	(0.004)	(0.004)	(0.004)	(0.004)
age	31.96	34.19	36.53	38.2
	(0.075)	(0.075)	(0.075)	(0.075)
age at first child	19.69	19.96	19.98	20.09
	(0.039)	(0.039)	(0.039)	(0.039)
works in agriculture	0.037	0.062	0.062	0.04
	(0.002)	(0.002)	(0.002)	(0.002)
works in sales	0.051	0.111	0.3	0.485
	(0.004)	(0.004)	(0.004)	(0.004)
works in services	0.771	0.474	0.284	0.09
	(0.004)	(0.004)	(0.004)	(0.004)
works in skilled & manual job	0.018	0.038	0.043	0.041
	(0.002)	(0.002)	(0.002)	(0.002)
works in technical & managerial job	0.016	0.071	0.067	0.115
	(0.002)	(0.002)	(0.002)	(0.002)
works in unskilled & manual job	0.008	0.021	0.024	0.022
	(0.001)	(0.001)	(0.001)	(0.001)
total years of education	7.277	7.646	7.654	8.757
	(0.039)	(0.039)	(0.039)	(0.039)
age difference to partner	6.593	4.729	3.831	2.35
	(0.059)	(0.059)	(0.059)	(0.059)
was minor at first sexual intercourse	0.933	0.763	0.695	0.57
	(0.004)	(0.004)	(0.004)	(0.004)

Table 41: Attrition bias: Values of covariates across the quartiles of the treatment effect

The quartiles are based on the predicted treatment effect and the values indicate the average value of the covariate in the quartile. The colors indicate the position of the mean of the subgroup in the standardized empirical distribution. The standardized distribution is colored from a scale of +/- 0.9. The causal forest is estimated using 0.8 of the sample as training sample with input parameters tuned once on the training sample and the domestic violence sample weight provided by DHS. The number of trees equals the number of women in the training sample (N = 46,153). The sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates, except the age difference in relationship. In the sample, 46,922 women indicated the value of the variable. I use this sample to predict the value of the variable for the 10,770 women with a missing value using a regression forest. The regression forest takes all main covariates except the age difference in the relationship to predict the value of the age difference in the relationship. The regression forest uses the domestic violence sample weights (DHS), min.node.size is set to 1000 and all other parameters are tuned. Less severe violence indicates whether the woman has ever experienced at least one of the following items: pushed, shook or had something thrown; slapped; punched with fist or hit by something harmful. The outcome only refers to actions ever committed by the husband or partner. The variable "experienced domestic violence as child" refers to actions committed by the father against the mother during childhood. "Minor" refers to all women younger than 18 years.

Variable	Relative Importance
age difference to partner	0.17
age at first child	0.15
age	0.14
total years of education	0.13
wealth quintile of household	0.10
was minor at first sexual intercourse	0.07
works in services	0.06
experienced domestic violence as child	0.06
works in sales	0.06
lives in urban environment	0.06
works in agriculture	0.00
works in technical & managerial job	0.00
works in skilled & manual job	0.00
works in unskilled & manual job	0.00

Table 42: Attrition bias: Sorted measure of variable importance

The measure of variable importance indicates how often a variable was used as splitting variable. The causal forest is estimated using 0.8 of the sample as training sample with input parameters tuned once on the training sample and the domestic violence sample weight provided by DHS. The number of trees equals the number of women in the training sample (N = 46,153). The sample is restricted to women with at least one living son who have ever been in a relationship/union/marriage, never had a multiple birth and have non-missing values for the outcome variables and the main covariates, except the age difference in relationship. In the sample, 46,922 women indicated the value of the variable. I use this sample to predict the value of the variable for the 10,770 women with a missing value using a regression forest. The regression forest takes all main covariates except the age difference in the relationship to predict the value of the age difference in the relationship. The regression forest uses the domestic violence sample weights (DHS), min.node.size is set to 1000 and all other parameters are tuned. Less severe violence indicates whether the woman has ever experienced at least one of the following items: pushed, shook or had something thrown; slapped; punched with fist or hit by something harmful. The outcome only refers to actions ever committed by the husband or partner. The variable "experienced domestic violence as child" refers to actions committed by the father against the mother during childhood. "Minor" refers to all women younger than 18 years.