CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

How Parental Involvement Affects Education Outcomes of Their Children

Master's thesis

Author: Bc. Klára Kantová

Study program: Economics and Finance

Supervisor: Mgr. Barbara Pertold-Gebicka M.A., Ph.D.

Year of defense: 2021

Declaration of Authorship

The author hereby declares that she compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain any other academic title.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis in whole or in part and agrees with the thesis being used for study and scientific purposes.

Prague, May 3, 2021

Klara Kantova

Abstract

Spending time together is more than important for family's well-being, especially in the fast pace of modern days. This Master's thesis presents a research of the relationship between parental involvement in children's lives and their educational outcome. Additionally, it explores the impact of limit breaking by youths on their educational attainment. The educational outcome is represented by a binary variable denoting whether the respondent completed high school or not. The results mostly meet our expectations. The hypotheses of the positive effect on child's educational outcome with higher parental involvement and negative effect with presence of limit breaking are supported by the results. What is surprising are the signs of the results from the regressions using the limit setting variables. The results suggest that the expected probability of completing high school decreases with higher parental limit setting. To estimate causal treatment effects, we used a subclassification on the propensity score and a simple logistic regression.

JEL Classification	I21, J12, J13
Keywords	education, children, family, parental involvement
Title	How Parental Involvement Affects Education Outcomes of Their Children

Abstrakt

Společné trávení času je pro blaho rodiny více než důležité, zejména v dnešní rychlé moderní době. Tato diplomová práce zkoumá vztah mezi angažovaností rodičů do života svých dětí a výsledky dětí dosažené na střední škole. Dále zkoumá dopad překročení limitu mládeží na jejich dosažené vzdělání. Výsledek vzdělávání je reprezentován binární proměnnou, která označuje, zda respondent dokončil střední školu či nikoli. Výsledky většinou splňují naše očekávání. Výsledky podporují hypotézu pozitivního vlivu na výsledek vzdělání dítěte s vyšším zapojením rodičů a negativního vlivu, když dítě překročí limit. Překvapivé jsou výsledky z regresí, které používají proměnné pro nastavení limitů rodiči. Výsledky naznačují, že očekávaná pravděpodobnost dokončení střední školy klesá, když má dítě vyšší počet limitů. K odhadu kauzálních efektů zapojení rodičů jsme použili párování pomocí propensity skóre a logistickou regresi.

Klasifikace JEL I21, J12, J13
Klíčová slova vzdělání, děti, rodina, angažovanost rodičů
Název práce Jak zapojení rodičů ovlivňuje vzdělanost jejich dětí

Acknowledgments

The author would like to express her great gratitude to Mgr. Barbara Pertold-Gebicka M.A., Ph.D.for her guidance, expertise and useful comments during conducting and writing this thesis. The author also wants to thanks doc. PhDr. Tomáš Havránek Ph.D. and the others for all their comments and help during the Master's thesis seminar. Finally, the author would like to thanks her family and friends for their support.

This thesis is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 681228.

Typeset in LaTeXusing the IES Thesis Template.

Bibliographic Record
Kantová, Klára: *How Parental Involvement Affects Education Outcomes of Their Children.* Master's thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2021, pages 94. Advisor: Mgr. Barbara Pertold-Gebicka M.A., Ph.D.

Contents

List of Tables

List of Figures

Acronyms

ATE Average Treatment Effect

ATT Average Treatment Effect on the Treated

ATUS American Time Use Survey

CI confidence interval

GPA Grade-point average

MMWS Marginal Mean Weighting through Stratification

MSE Mean Squared Error

NLSY97 National Longitudinal Survey of Youth 1997

OR odds ratio

PSM Propensity Score Matching

SD standard deviation

SMD Standardized Mean Difference

Master's Thesis Proposal

Author	Bc. Klára Kantová
Supervisor	Mgr. Barbara Pertold-Gebicka M.A., Ph.D.
Proposed topic	How Parental Involvement Affects Education Outcomes of Their Children

Motivation Spending time together is more than important for family's well-being, especially in the fast pace of modern days. Nowadays, it is easily forgotten that spending time with children is the most precious and really important thing. It is proved that spending time with family has positive impact on children. It builds children's self-esteem, strengthens family bonds, develops positive behaviors, encourages communication, and – last but not least – it can help child's academic performance. There is significant correlation between parental education and parental time with children as well. Focusing on the mothers, one would expect that higher maternal education would lead to less time spent with children, meaning that they want to go back to work and they want to end their maternity leave faster than less educated mothers. Moreover, they are wealthier and can afford au-pair with higher probability than less educated mothers. However, studies show that higher-educated parents actually spend more time with their children. Guryan, Hurst, and Kearny (2008) found that mothers with a college education or higher spend more than 4 hours per week more with their children than mothers with lower education. Also, Kalil et al. (2012) tested the hypothesis that highly educated mothers spend more time in active childcare than mothers with lower education. In my diploma thesis, I would like to focus on how parents spend time with their children and how this affects children's education later in life. Institute of Education (2002) claims that parental involvement in child's early years has significant effect on their cognitive development, literacy and number skills. Moreover, Feinstein & Symons (1999) claim that parental involvement in a child's schooling between the ages of 7 and 16 is more powerful than family background, size of a family, or level of parental education. Also, educational failure is increased by lack of parental interest in schooling.

Hypotheses

Hypothesis #1: How parental involvement affects their childrenâ€™s education?

Hypothesis #2: Is it important HOW do parents spent their time with their children (in terms of children's education)? Or is it enough, that they are spending time together, no matter how?

Hypothesis #3: Is father's interest in a child's schooling more important than mother's interest?

Methodology To test the above-mentioned hypotheses, construction of a new panel data set has to be made. It will be done so by the combination of American Time Use Data (ATUS), from where I would obtain information about time use measures, and the National Longitudinal Survey of Youth 1997 (NLSY97), where the education of children is recorded. The merge will be done by Ginga (2010), where it is clearly discussed how to combine three data sets (ATUS, NLSY79, and CEX) into a panel. The econometric model will have child's educational outcome as a dependent variable. Dependent variable would be measured when children are 18 years old by high school GPA. The main explanatory variable would be parental time spent with their children. I will work with four measures of the activity (Basic care, Play, Teaching, and Management) in age of children 5-12. Not adding parental education could cause endogeneity in our model as the child's educational outcome and also parental time spent with their children are both dependent on parental education. However, the information about parental education is hard to obtain. Moreover, there is also potential spurious correlation because there are other factors that correlate with both, time spent with children and children's educational outcome, e.g. there is correlation between parental IQ and child's IQ, highly educated parents are usually wealthier and can afford better schools, etc. To solve endogeneity in our model, we introduce matching method by matching similar families and compare two children from those families. Estimation of the impact of parental involvement on children's education will be made by the described methodology. Also, I would like to focus on the difference in paternal and maternal involvement in a child care.

Expected Contribution Existing literature mostly focuses on younger children or uses different data. Neidell (2000) investigates early parental involvement (in the first year of child) on cognitive and non-cognitive outcomes. Rasmussen (2009), and Price (2010) both deal with the impact of parental involvement on child's educational outcome. However, they both work with different data. I will combine ATUS data and NLSY97 data to estimate the impact of how parents spend time with their children on children's education. In contrast to the existing literature, I will focus on exact activities how parents spend their time with them. In addition, maternal and paternal involvement will be compared. The aim of the Thesis is to show that parental time with their children is precious (not only) in terms of future children's education. Moreover, I wonder if there is higher positive correlation in paternal time spent than maternal in children's education.

Outline

1. Introduction
2. Literature review
3. Data
4. Methodology
5. Results
6. Conclusion

Core bibliography

Desforges, C., 2003. The Impact of Parental Involvement, Parental Support and Family Education on Pupil Achievement and Adjustment. DfES Research Report 433.

Feinstein, L., Symons, J., 1999. Attainment in Secondary School. Oxford Economic Papers.

Ginga, R., 2010. Income Shocks and Investments in Human Capital. University College London.

Griffith, J., 1996. Relation of parental involvement, empowerment, and school traits to student academic performance. The Journal of Educational Research, 91 (1), p. 33-41.

Guryan, J., Hurst, E., Kearney, M., 2008. Parental Education and Parental Time with Children. Journal of Economic Perspectives, Volume 22, Number 3, p. 23-46.

Kalil, A., Ryan R., Corey M., 2012. Diverging Destinies: Maternal Education and the Developmental Gradient in Time with Children. Demography (2012) 49, p. 1361-1383.

Neidell, M. J., 2000. Early Parental Time Investments in Children's Human Capital Development: Effects of Time in the First Year on Cognitive and Non-cognitive Outcomes. UCLA Department of Economics.

Porterfield, L. S., Winkler, E. A., 2007. Teen Employment: Shifting Patterns by Parental Education and Family Structure. Monthly Labor Review, p. 37-56.

Price, J. 2010. The Effect of Parental Time Investments: Evidence from Natural Within-Family Variation. Department of Economics, Brigham Young University and NBER.

Rasmussen, W. A., 2009. Allocation of Parental Time and the Long-Term Effect on Children's Education. Department of Economics and CIM, Aarhus University.

Chapter 1

Introduction

Spending time together is more than important for family's well-being, especially in the fast pace of modern days. Nowadays, one can easily forget that spending time with children is the most important thing. There is no doubt that spending time with family has positive impact on children. There are various sectors affected by the time parents spend with their children. It builds self-esteem of children, strengthens family bonds, develops their positive behavior, and encourages communication. Finally and most importantly for this thesis, it can affect child's academic performance.

The subjects of this thesis are Americans born between 1983-1984 with both birth parents present in the household when the respondents are 12 to 14 years old. This limitation is given by the characteristics of the explanatory variables used. The explanatory variables are the variables designing the level of parental involvement. The level of parental involvement is measured by asking youths and their parents questions relating to monitoring, limit setting, and limit breaking. Respondents and their parents were asked those questions when respondents were 12 to 14 years old. Data are collected from the National Longitudinal Survey of Youth 1997, which is publicly available database.

The objective of this thesis is to research on the impact of parental involvement on the educational outcome of their children. It focuses on the parental involvement in child's everyday life and how it affects his or hers educational outcome later in life. As an educational outcome we provide a binary variable determining whether the respondent completed high school or not. As the explanatory variables we present a vector of six parental involvement variables.

According to Guryan *et al.* (2008) and Kalil *et al.* (2012), higher educated parents spend more time with their children than less educated parents. Moreover, McLanahan (2004) argues that children of more educated women are gaining more assets such as parental time and money. From that follows that more educated parents spend more time with their children, however, we do not know whether it affects child's educational outcome. Our goal is to investigate whether parental involvement which is correlated with parental level of education based on mentioned studies affects child's education attainment, i.e., graduation from high school. We expect that the effect is greater with paternal involvement than with maternal. Moreover, we investigate the effect of parental involvement on child's educational outcome with the presence of variables which control the characteristics of children as well as characteristics of their parents. We select a vector of control variables based on the findings in the literature. In the literature, we found that gender, race, household size, location, and parental level of education may be correlated with the educational outcome of children.

We provide the empirical results from simple data observing, simple logistic regression, and from subclassification matching method. We regress a binary variable determining whether the respondent completed high school or not on a vector of parental involvement variables. In addition, we include a vector of control variables to control the differences in demographics which may affect not only the probability of graduation from high school but also the variables of parental involvement, i.e., explanatory variables.

Our key finding is that parental involvement depicted by monitoring variables positively relates to the child's educational outcome. We can interpret this as that parents should be involved in child's life in order to improve children education performance. Additionally, our results indicate negative impact of limit breaking by children on their education outcome. Higher effect of paternal involvement is not supported by our findings. Moreover, looking at the limit setting variables, our findings indicate that limit setting is negatively related to the child's educational outcome. Our results suggest that higher limit setting by parents results in a reduction in the probability of graduating from high school. This may be caused by our limited sample or by that we do not control for the "problematicity" of children and family's wealth.

There are studies concerning the relationship of education level of parents and parental time spent with children, however, we want to go further. The studies mainly focus on parental engagement in the terms of school - having discussion about the school, helping with homework, reading with children, and concerning also parental involvement at school such as parents' volunteering, attending the workshops, school plays, sport events, school political involvement etc. Very few authors focus on the effect of HOW parents interact with their children and children's education outcomes.

In this thesis, we focus on parental involvement in everyday life. This study contributes to existing literature in various ways. First, this is the first study which examined the effect of parental time spent with children on their educational level in terms of the engagement of parents in child's everyday life. Second, this work contributes to a gap in literature by including the characteristics of children as well as parents. In conclusion, our study assists in better understanding of the role of parents in educational outcomes of their children.

The remainder of this thesis is structured as follows: Chapter 2 contains a literature review. Chapter 3 gives an introduction to data used for the analysis. It includes detailed description of used variables and it covers basic demographics and descriptive statistics of our sample. Chapter 4 describes a methodology used in this thesis. Chapter 5 presents results and discussion. Finally, Chapter 6 concludes our findings.

Chapter 2

Literature review

Raising a child requires investing material resources as well as time. Parental time spent with their children is the best investment in extending their human capital. As this thesis is focused on parental involvement and its effect on children educational outcome, the literature review focuses on the investment in parental time spent with children rather than material investment.

Mcdowell *et al.* (2018) claim that parental involvement in child's early years has significant effect on their cognitive development, literacy, and number skills. Moreover, Feinstein & Symons (1999) claim that parental involvement in a child's schooling between the ages of 7 and 16 is more powerful than family background, size of a family, or level of parental education in terms of secondary school achievement. Also, child's educational failure is increased by lack of parental interest in schooling.

2.1 Educated parents spend more time with a child-care

Even though Feinstein & Symons (1999) claim that parental involvement is more powerful than parental education, many studies show that parental education level is correlated with the amount of time spent with children. Studies argue that higher-educated parents actually spend more time with their children. Guryan *et al.* (2008) found that mothers with a college education or higher spend more than 4 hours per week more with their children than mothers with lower education. Their study examines 4 groups – working mothers, non-working mothers, working fathers, and non-working fathers, the relationship found holds for all groups. This finding is very surprising as more educated parents spend more time in a work (outside the home). The result not only holds among all examined groups, but also across all four examined childcare activities – basic, educational, recreational, and travel activities related to a child. Moreover, why are these results striking is that opportunity cost of time is much higher for high educated parents (i.e., higher waged according to the author) than for less educated.

A study by Guryan *et al.* (2008) also examines whether the observed relationship of parental education and time spent with children in the US holds in other countries too. It examines 13 countries (e.g., Norway, UK, Netherlands, Canada, Chile, South Africa, Palestine, . . .) and compares the results with the US. The sample is restricted to the households which includes individuals (adults, mothers, and fathers) ages of 21 to 55 inclusive and have at least one child younger than 18 in the household. For almost all observed countries, the same result holds as for the US. Meaning that more educated adults, thus probably adults with higher income, spend more time in a childcare than less educated adults, probably with lower income. Moreover, it follows that time in a childcare is much more appreciated by higher educated individuals. In other words, by adults with higher opportunity cost of time. As the outlier can be considered Chile since it is the only country where the difference in the impact of higher and less educated adults on time spent in a childcare is not significantly different from zero.

Kalil *et al.* (2012) tested the hypothesis that highly educated mothers spend more time in an active childcare than mothers with lower education. Their study not only examines that high educated mother spends more time with her child, but also whether the distribution of her time spent in a childcare is more effective than by less educated mother. The largest differences were found in the youngest (0-2 y.o.) age group. The results support the hypothesis that more educated mothers spend more time with their children than less educated. Specifically, mothers with a high school degree of children in ages 0-2 spend about 67 minutes per day less with their children than college (or beyond) educated mothers. Moreover, mothers with a high school degree of children in ages 3 to 5 and 6 to 13 spend with their children 21 minutes and 22 minutes per day less, respectively, than college (or beyond) educated mothers. Concerning activity type, more educated mothers distribute their time more effectively.

Lastly, McLanahan (2004) argues that children of high educated women are gaining assets such as parental time and money. Thus, on the other side, children of low educated women are losing those. This study was conducted with a purpose that government should focus on increasing inequality between rich and poor children and that government should implement some measures to close this widening gap.

As we can see, the literature mostly focuses on women in a childcare, even though the gap between maternal and paternal time spent in a childcare is closing as is mentioned in the next subchapter. Thus, the literature should focus on both women and men.

2.2 Households with both birth parents

Paper by Rasmussen (2009) which focuses on families with two full-time employed parents analyzes, theoretically and empirically, relationship between parental time use and child development. Moreover, it investigates the differences in quality of market-provided childcare and quality of parental childcare. The impact of parental involvement on children's educational outcome is monitored in Denmark using Danish time use data. The author finds statistically significant relationship between mothers' time spent with children during weekdays and child's educational outcome. What is more, the author finds statistically significant results for the relationship between fathers' childcare time on weekends and children's educational outcomes. Significant negative relationship is observed between parental time spent with children and working time.

The reason why this study focuses on households with both parents present is increasing number of recent studies about distribution of housework and childcare between a mother and a father. In 2008, a study about housework trends is released as a part of the study on income dynamics. The study is comparing the years 1976 and 2005. Generally, the number of hours spent on chores decreased for women by almost 10 hours per week (to 26 hours per week) during the reference period and doubled for men up to 12.5 hours over the same period. Meaning that the gap between mothers' and fathers' hours spent on housework is closing. What is more, the activities such as home repairs, shoveling snow, or mowing the lawn which are mostly done by men were not included in the research at all. When including these, the gap is supposed to close even more. A study conducted by American Time Use Survey (ATUS) compares average minutes per day men and women spent in household activities in 2015. The activities are distributed into following sections: food and drink preparation, interior cleaning, laundry, household management, other household activities, kitchen and food clean-up, lawn garden and houseplants, maintenance repair and decoration. The order of activities is the time women spent on the most to activities on which women spent the least time. Overall, women spent on those household activities 2 hours 15 minutes per day on average and men 1 hour 25 minutes. Speaking of caring for and helping household members, women spent twice (0.6 hour per day on average) as much time as men (0.3 hour per day on average) by those activities. Concerning specifically childcare in households with children under the age of 6, women spent on average 1 hour per day by giving physical care (bathing or feeding). Whereas men spent only 25 minutes per day by giving physical care. Those data are related with non-institutional people aged 15 or older in 2015.

2.2.1 Parental time distribution

During different stages of child's life, parents distribute their time in a childcare differently as in every stage it is important to focus on development of different qualities.

Infants (from birth to 12 months) described in Bornstein (2005) require mostly basic care such as comforting, feeding, and bathing. Parents invest their time in establishing eating and sleeping routines. Toddlers (from 12 to 35 months) acquire basic skills such as cognitive and social skills, they require attention and develop logical reasoning, memory, vocabulary, but also emotional and behavioral qualities. This is achieved by playing with a toddler so parents should focus on directed child play during toddlerhood. In preschool period (from 3 to 5 years), the importance is still focused on directed plays but more emphasis should be placed on teaching activities such as reading books and solving any kinds of "problems" (e.g., puzzles, etc.). Snow (2006) claims that teaching activities are very important in a preschool period, because it affects early educational outcomes of children. In middle childhood (from 6 to 13 years) children start to evolve more outside of the family. They start to attend school and extracurricular activities. From this stage onward, parents should actively manage, plan, and monitor child's academic and social networks.

2.3 Importance of setting limits with children

Setting limits to the child from the early age helps to develop his or hers discipline. Gained and developed discipline then improve child's future academic performance.

According to Morin (2021) there is a difference between setting the limits and rules. The limits express the guideline for behavior and give opportunities to deepen children's skills. If they are established at an early age of the child, it will make later education easier and it becomes a habit. The author mentioned five main reasons why limit setting is important. The limits teach children self-discipline, keep them safe and healthy. Moreover, they help them cope with uncomfortable feelings and show them that their parents care.

Self-discipline skills are acquired by setting the limits. It is a part of being responsible. When parents tell the children that they should do their homework and shut the video game down, they are trying to teach them the habit of responsibility. The goal is to teach them to manage all responsibilities, such as homework, chores, brushing teeth, etc., without reminding them to do so. According to Morin (2021), simple way how to teach children self-discipline from an early age is telling them phrases as "you can have sweets after you finish your meat and vegetables", or parent tells child to "beat the timer" when he or she is getting ready in the morning, or " you can watch TV after you finish your homework".

There is no doubt that the limits keep children safe, however, parents ban some activities instead of setting the limits. Setting the limit instead of banning some activity is teaching children how to keep themselves safe. Moreover, parents are giving opportunity to the child to prove that he or she can be responsible with those limits and that he or she may be able to handle more responsibility.

The limits keep children healthy. Not only in the terms of eating habit, but also, in terms of sport, hygiene, etc. As we already mentioned, it is good to set some limits on eating habits (sweets after the healthy part of the meal, etc.). Without the limits, many children would always pick junk food rather than the healthy choice. To be sure that children would follow a healthy lifestyle, it is good to set the limits also with screen time, exercise, and hygiene.

Many parents avoid setting the limits as they do not want to make uncomfortable their children as well as themselves. However, the limits help children to handle uncomfortable feelings, which is an important skill for the future. For the parents, it is important to not to feel guilty when their child feels frustration, boredom, anger, or sadness because of the limit. With each limit comes the opportunity for the child to try to manage the emotions.

As the last point for setting the limits, we would like to stress that according to Morin (2021), children often like to test parents how serious they are about the limit. However, that does not mean they do not want to have those limits. The limits are giving to children the feeling of parental care. Children with a few limits or even completely without the rules experience anxiety.

Gordon (2017) distinguishes four types of limits which should be set by parents to their child. The first is the safety limit. The limit which should stop a child from breaking something, hurting someone or even hurting himself or herself. Secondly, the author talks about value limit. Value limit regards to upholding family values, aims, and traditions. The third is the expectation limit. The expectation limit encourages children that they are good enough to fulfill a request or to try something even when their fears hold them back. Lastly, we have proposal limits. These are mainly concerned about good byes. They have their place here because by proposal limits parents are trying to avoid the unknown situations and to bring child's feelings to the surface.

2.3.1 Consequences

It is important to think about consequences as well. Not only about the negative consequences when the limit (or rule) is broken but also about the positive ones. The negative consequences should surely be a part of the upbringing of children, however, we are convinced that positive consequences are as important as negative. Morin (2020) claims that positive and negative consequences should be used together and if they are used together properly as effective discipline tool, they will change child's behavior. The author advises to strengthen good behavior by using positive consequences, and oppositely, to discourage bad behavior by enforcing negative ones.

According to Morin (2020), parents should start teaching their children the impact of breaking the limits or fulfilling the requests (i.e., consequences) from an early age. They can learn from childhood that good choices as doing chores, homework, etc., lead to positive consequences. Oppositely, bad choices, as is physical aggression, lead to negative consequences. The most effective consequences are those which are immediate and consistent.

2.4 Importance of high school in the US

According to Zaff *et al.* (2017) nearly one in five students does not complete high school (i.e., does not graduate) on time, if ever. Completing high school subconsciously teaches the students how to positively contribute to economy as well as to civic life.

As claimed by Belfield & Levin (2007), graduation from high school is a doorway to college (or different higher education), economic self-sufficiency, and civic engagement. Without a high school diploma, people are more likely to earn lower income and to be arrested, which leads to higher costs for the US. According to Sum *et al.* (2009), each high school dropout costs the US about $292,000 more than the high school graduate (over lifetime). It is caused by lower taxable income together with higher reliance on social welfare programs.

Zaff *et al.* (2017) claim that high school dropouts are associated with low-income families and neighborhoods, and also with ethnicity. According to Zaff *et al.* (2017), only 72 % students from low-income families graduate from high school on time. On the other side, 87 % students from high or middle income families graduate from high school on time. Concerning the ethnicity, African-American and Hispanic students have about 10 % lower graduation rate than the national average. Children's graduation is associated with their parents' education. There is a difference in health, economic, and educational outcomes of children with those parents who have graduated from high school and those who have not.

McCallumore & Sparapani (2010) examine the importance of ninth grade on high school graduation. According to their study, the increase of high school dropouts in 2001 occurred when there was a significant emphasis on obtaining college degree. Students felt under the pressure and did not even complete high school. The focus of the study is on the ninth grade because it claims that the problem of dropouts arises from the transition from middle school to high school. This transition affects the graduation. They called the ninth grade as "make or break year". As a solution, they introduce reform programs such as freshman academics which are supposed to help ease the transition from middle school to ninth grade. However, they also pointed out that the problem of high school dropouts does not arise here (in the transition), but it starts in early age of the students. In the kindergarten, students do not learn everything they should, nor are they in the middle school. What is more, middle school teachers do not cooperate with high school teachers. Thus, the graduation from high school is of a significant concern in the US.

2.5 Parental involvement in child's academics

Most studies dealing with the topic of the impact of parental involvement on children's educational outcomes investigate the parental involvement in child's academic performance such as having discussion about the school, helping with homework, reading with children, and include also parental involvement at school such as parents' volunteering, attending the workshops, school plays, sport events, or school political involvement.

According to Cole (2017), the more parents are involved in the education of their children, the more students are likely to excel in academic performance. Thus, they more likely become a productive members of the society. Also Bryan (2005) claims that if parents are actively participating in children's education, they are more likely to excel in academics.

Hoover-Dempsey & Sandler (1997) looked at the issue from a different perspective. They studied why parents become involved in their children's education from a psychological point of view. They ended up with three main areas which affect parents' involvement decisions. First, the construction of the role of parents defines parents' beliefs about what to do in their children's education and it seems to set the basic scope of activities that parents consider to be important. Second, parents know that helping their children to success in a school is effective. They focus on the extent to which they believe that through their involvement they can have a positive effect on their children's educational outcomes. Third, their child and also the school, they want the parents to be involved.

2.6 Summary

From this chapter follows several important points. First, more educated parents spend more time with their children. Moreover, they distribute their time with children more effectively. Second, we should focus on increasing inequality between rich and poor children. Rich parents usually live in better and safer locations and have access to better schools than poor ones. Those facts can also affect the educational outcome of children. Third, gained discipline through limits setting improve child's academic performance. Children with a few limits or even completely without the rules experience anxiety. Finally, graduation from high school improves economic self-sufficiency and civic engagement. Without a high school diploma, people are more likely to earn lower income and to be arrested, which leads to higher costs for the US.

Parental involvement certainly impacts student academics, however, the examined studies involve parental involvement in terms of participation in the educational process of the child. We examine the effect of parental involvement in child's everyday life on child's educational outcome.

Chapter 3

Data

The main data for this study comes from the National Longitudinal Survey of Youth 1997 (NLSY97). NLSY97 is a longitudinal project that follows the lives of a sample of Americans born between 1980-1984. This survey is a part of the National Longitudinal Surveys program. Data observes 8,984 individuals over time. Respondents' ages ranged from 12 to 18 when first interviewed in 1997-1998. Now, there are available 18 rounds of interviews with the last round, round 18, held in 2017-2018, when respondents were 32-38 years old.

In this chapter, we precisely describe the data and our sample used in the analysis. We present the dependent variable, explanatory variables, and we introduce a vector of control variables. Finally, we present the correlation between education attainment variables of parents and respondents.

3.1 Sample

The main research question relates to the relationship between parental involvement and children's educational outcomes. The questions about parenting techniques (i.e., parental involvement) were only asked of those respondents whose age ranged from 12 to 14 years when first interviewed. So we limit our sample to only these respondents. In the data, 3,578 respondents fulfill this criterion. However, not all of these respondents answered the questions concerning either education attainment or parental involvement. When we omit all NA (not available) responses, the sample of 3,578 observations is then reduced to 1,868 observations. So the original dataset containing 8,984 observations is now reduced to the sample of 1,868 observations. We will use this reduced sample for further analysis.

In Table 3.1 we can see the comparison of the two samples. Both include respondents who are born in either 1983 or 1984. In the first column, we depict the sample of 1,868 observations where we omitted all NA values. In the second column, we derive the sample of 3,578 observations where NA values are included. The first four rows present categorical variables with the corresponding number of respondents in each category. Moreover, we denote percentages of the total in parentheses. Then, we present three variables with their means and standard deviations in parentheses. In addition, we denote the ranges of the values for better understanding. Finally, in Table 3.2 we present t-tests.

Table 3.1: Comparison of statistics

		sample of 1,868 obs.	sample of 3,578 obs.
Gender	Male	976 (52.2 %)	1845 (51.6 %)
	Female	892 (47.8 %)	1733 (48.4 %)
Race	Black	297 (16.0 %)	904 (25.3 %)
	Hispanic	359 (19.2 %)	760 (21.2 %)
	Mixed	10 (0.5 %)	32 (0.9 %)
	White	1202 (64.3 %)	1882 (52.6 %)
Location	Urban	1290 (69.0 %)	2619 (73.2 %)
	Rural	498 (26.7 %)	800 (22.4 %)
	Unknown	80 (4.3 %)	159 (4.4 %)
HS	Yes	1576 (84.4 %)	2743 (76.7 %)
	No	292 (15.6 %)	822 (23.0 %)
	NA	0	13 (0.3 %)
HH size	Mean	4.72	4.61
		(1.30)	(1.51)
	Range	<2; 14>	<1; 16>
hgc M	Mean	12.89	12.51
		(2.98)	(2.96)
	Range	<2; 20>	<0; 20>
hgc F	Mean	12.85	12.52
		(3.27)	(3.26)
	Range	<2; 20>	<0; 20>

Note: HS - the information about whether the respondent completed high school or not; HH size - household size; hgc M - the highest degree completed by mother; hgc F - the highest degree completed by father.

Concerning the categorical variables, we can conclude that the portions of males and females remain almost the same. Distribution of location variable also remain almost the same. Speaking of race, we observe significant differences for Black and White respondents. The portion of respondents who graduated from high school increases with the reduced sample.

Moreover, we can observe slight changes in the means. We can see that the mean for household size slightly rises and the standard deviation slightly declines after omitting the NA values. The means of the highest grade completed by both mothers and fathers rise as well. However, the standard deviations remain almost the same.

Table 3.2: Welch Two Sample t-test

	Test	Results
hgc M	Welch Two Sample t-test:	t(3850.60) = -4.50, p < .001, d = NA
hgc F	Welch Two Sample t-test:	t(3977.87) = -3.41, p = .001, d = NA
HH size	Welch Two Sample t-test:	t(4287.03) = -2.79, p = .005, d = NA

Note: hgc M - highest degree completed by mother; hgc F - highest degree completed by father; HH size - household size.

In the above mentioned table, we can see the results from the t-tests. The Welch t-test is performed to observe the mean difference between the samples. The p-values are lower than the significance level 0.05, so we reject the null hypothesis meaning that significant differences were detected. In other words, we can conclude that the mean values of the three observed variables are significantly different across the samples. However, we use the smaller sample of 1,868 observations since it includes all the information we need for the analysis.

Chosen variables for high school graduation as well as for parenting techniques are described in detail in the following subchapters. We start with presenting the dependent variable, i.e., high school graduation variable. Then we present the explanatory variables. Additionally, we introduce the control variables.

3.2 Characteristics

3.2.1 Data description – dependent variable

The NLSY97 data contains information about educational outcomes of the respondents.[1] For the purpose of our study, we use the information about the highest degree completed by the respondents as the dependent variable. The variable for the highest degree completed is a cumulative variable. This vari-

[1] The information about their GPA is calculated by the school in its metric for the last year of youth's enrollment. As the metric varies significantly across the US countries, we cannot use this information as the educational outcome since those values are incomparable.

able is created for each respondent in a different point of time. It is collected after each respondent finishes the last school.

From our sample, 1,576 (84.4 %) respondents have received a high school diploma (have completed regular 12 years program). Higher education than that is completed by 804 (43 %) individuals, from which 173 respondents completed Associate/junior College (AA), 430 respondents completed Bachelor's degree (BA, BS), 147 respondents completed Master's degree, 12 respondents completed PhD program, and 42 respondents completed Professional degree (DDS, JD, MD). This is summarized in Table 3.3.

Table 3.3: Highest degree received by children

	n	%
None	137	7.3 %
GED	155	8.3 %
High school diploma	772	41.3 %
Associate/Junior college	173	9.3 %
Bachelor's degree	430	23.0 %
Master's degree	147	7.9 %
PhD	12	0.6 %
Professional degree	42	2.2 %
	1,868	100 %*

* not equal to exactly 100 % due to rounding

To estimate the effect of parental involvement on child's educational outcome, we have to define an appropriate outcome variable. The variable of the highest degree received by children is a categorical variable with 8 different values, which would be difficult (impossible) to interpret results with. Thus, we introduce a dummy variable whether the respondent completed high school or not. In the US, the most important part of education is the graduation from high school as discussed in the Chapter 2.

In Table 3.4 we can see the portion of those who have completed high school and those who have not. We can see that the sample is unbalanced, meaning that there are much more respondents who have graduated from high school than those who have not. Moreover, we can observe the portions also by gender. We can conclude that the portion of males who have completed high school to the total males is lower than for females.

Table 3.4: Graduation from high school

HS	n	%	male	%	female	%
YES	1576	84.4%	796	81.6%	780	87.4%
NO	292	15.6%	180	18.4%	112	12.6%
	1,868	100%	976	100%	892	100%

Note: HS - the information about whether the respondent completed high school or not.

3.2.2 Data description – explanatory variables

For the purpose of our study, we choose variables from NLSY97 such as monitoring youth by mother as well as by father, limit setting and limit breaking reported by youth as well as by parent.

Monitoring

Information about monitoring is obtained by asking the respondents four questions. The four questions are:

(i) How much does he or she know about your close friends, that is, who they are?

(ii) How much does he or she know about your close friends' parents, that is, who they are?

(iii) How much does he or she know about who you are with when you are not at home?

(iv) How much does he or she know about who your teachers are and what you are doing in school?

Responses to these questions are measured on a 5-point scale (0 - parent knows nothing, 1 - knows just a little, 2 - knows some things, 3 - knows most things, 4 - knows everything). The questions are separated regarding mother and father, so eventually, there are two variables for monitoring. Monitoring youth by mother and monitoring youth by father. The parental monitoring scale is created for each of the four possible parental figures (residential mother, residential father, non-residential biological mother, non-residential biological father). However, this thesis focuses only on residential parents, so we only use the two of them. After the responses are collected, the monitoring variable[2] is

[2] Scale.

created by summing these responses, so the value ranges from 0 to 16, where higher score indicates greater parental monitoring. Buchanan *et al.* (1992) claim that degree of monitoring has impact on child's scholastic achievement, behavior, or even on sexual involvement (lower degree of parental monitoring is linked with lower education, behavioral problems, and early sexual involvement).

Autonomy, Control, and Limit setting

Information about autonomy, control, and limit setting is obtained by asking the respondents three questions. Parallel questions were asked of the parent. Responses to these questions are measured on a 3-point scale (0 - parent let youth decide, 1 - parents and youth decide jointly, 2 - parent or parents set limits).

(i) Who set the limits on how late you stay out at night? / how late he or she can stay out at night?

(ii) Who set the limits on who you can hang out with? / who he or she can hang out with?

(iii) Who set the limits on what kinds of TV shows or movies you can watch? / what kinds of TV shows and movies he or she can watch?

Then the youth limits setting index[3] is created by summing these responses. So the value of the limit setting variable responded by youth ranges from 0 to 6 where higher score indicates greater parental role in the limit setting. Similarly, the limit setting index responded by parents is created by summing up their responses. The range is the same as for youths – from 0 (youth sets all limits) to 6 (parent sets all limits).

According to Erford (1995), observing parallel questions is very useful as discrepancies across the answers indicate misunderstanding about who in fact sets the limits, which often leads to limits breaking by youths from the parents' point of view. However, this thesis does not treat uniquely those discrepancies.

[3]Setting a limit on one activity is not necessarily correlated with another. Thus index, not scale.

Limit breaking

Information about the limit breaking is obtained by asking the respondents three questions. Parallel questions were again asked of the parent. Responses to these questions are measured on a 3-point scale.

(i) In the past 30 days, how many times have you broken the limits about how late you can stay out at night? / how many times do you think he or she has broken the rules about how late he or she can stay out at night?

(ii) In the past 30 days, how many times have you broken the limits about who you can hang out with? / how many times do you think he or she has broken the rules about who he or she can hang out with?

(iii) In the past 30 days, how many times have you broken the limits about what kinds of TV shows and movies you watch? / how many times do you think he or she has broken the rules about what kinds of TV shows and movies he or she can watch?

Then the limits breaking index[4] responded by youths is either 0 - did not break the limits, 1 - broke any of three limits, or 9 - youth sets all three limits, meaning that there are no limits. Similarly, the limits breaking index responded by parents is 0 if parent reported that youth did not break any of these limits, 1 - parent thinks that youth broke any of limits, or 9 - parent reported that youth sets all three limits, thus there are no limits to break. Again, the discrepancies are observed here. About 13 % of youths claim that they did break at least one of the limits even though their parent reported that they did not break any of these three limits.

Creation of dummy variables

The variables for parenting techniques are designing the level of parental involvement in child's life. Hence, we treat them as factors. In Appendix A we can find table with descriptive statistics of parenting techniques for our reduced sample of 1,868 respondents.

Furthermore, for the purpose of our analysis, we create a dummy variable equal to 1 with high involvement of parents, and equal to 0 otherwise. The boundaries of the groups are set based on the level of parental involvement.

[4]Breaking a limit on one activity is not necessarily correlated with limit-breaking on another activity. Thus index, not scale.

Concerning the monitoring variables by mother and father, the values from 0 to 8 design low level of parental involvement, and values from 9 to 16 represent high level of parental involvement. The monitoring variables are denoted by summing the answers to the four aforementioned questions. The responses with values either 0, 1, or 2 (parent knows nothing, knows just a little, knows some things) represent low parental involvement, and the responses with values 3 and 4 (parent knows most things, parent knows everything) represent high parental involvement. Based on those facts, we set the boundary at value 8. Regarding to the limit setting variables reported by youths and parents, low level of parental involvement is represented by the values from 0 to 3. Higher values than 3 represent high level of parental involvement. The limit setting variables are denoted by summing the answers to the three previously mentioned questions. The responses are either 0, 1, or 2, parent let youth decide, parents and youth decide jointly, parent set limits, respectively. Thus, the upper bound for low parental involvement is set at the value 3. The variables showing if youth have broken any limit in the past 30 days reported by youths as well as by parents already were dummy variables. They are equal to 1 if children have broken any limit, and 0 otherwise. The value of 9, meaning that the youth sets all three limits, has not been reached by a single respondent in the sample.

In the following table we can see modified variables of the parenting techniques. Table 3.5 presents the number of respondents in each group after the modification. In Appendix B, we present histograms of the distribution of the parental involvement variables before the creation of dummies.

Table 3.5: Number of participates in treated and control group

Group	Mon M	Mon F	Limit Y	Limit P	Broke Y	Broke P
0	408	897	994	1050	1056	1311
1	1460	971	874	818	812	557

Note: Mon M - child monitored by mother; Mon F - child monitored by father; Limit Y - limit setting by parents, youth report; Limit P - limit setting by parents, parent report; Broke Y - youth broke limit, youth report; Broke P - youth broke limit, parent report.

3.3 Control variables

Even though Feinstein & Symons (1999) argue that parental involvement is more powerful in terms of succeeding in a secondary school than family background, size of a family, or level of parental education, there are number of studies as mentioned in Chapter 2 that show that parental education, neighborhood or household size affect child's educational outcome. Thus, we introduce a vector of control variables. The vector contains variables for gender, ethnicity, location (i.e., neighborhood), and education of both parents.

In Table 3.6 we can see categorical variables which are used in our analysis (gender, ethnicity, location). We denote the number of respondents in a particular group and percentage of the total.

Table 3.6: Categorical control variables

	n	%
Male	976	52.2%
Female	892	47.8%
Black	297	16.0%
Hispanic	359	19.2%
Mixed race (Non-Hispanic)	10	0.5%
Non-Black/Non-Hispanic	1202	64.3%
Urban	1,290	69.0%
Rural	498	26.7%
Unknown	80	4.3%

We can see that the portion of males and females is almost the same as it was before restricting the age limit and omitting the NA values of parenting techniques variable (i.e., as in the original dataset of 8,984 observations). Concerning the ethnicity, the portion of Non-Black/Non-Hispanic respondents increases from 51.9%, thus the other portions decrease. For Mixed race and Hispanic respondents the decline is not that high as for Black respondents. There is a decrease of 10% (from 26% to 16%). Regarding residential area, there is about 3% change, meaning that rural residents increase their portion by 3% and urban residents otherwise.

In Table 3.7 we can observe descriptive statistics of variables within our sample.

Table 3.7: Descriptive statistics

	Mean	SD	Median	Min	Max
HH size	4.72	1.30	4	2	14
hgc M	12.89	2.98	12	2	20
hgc F	12.85	3.27	12	2	20

Note: HH - Household; hgc F - Highest grade completed by father; hgc M - Highest grade completed by mother.

Speaking of household size, we can observe that the mean slightly rises from 4.55 to 4.72 compared to the original dataset. The median is exactly the same, and the range narrows as the minimum value increases from 1 to 2 and the maximum value drops from 16 to 14. The standard deviation decreases from 1.54 to 1.30. Concerning the highest degree completed by parents, we can conclude that both means rise, the medians remain the same, and the range narrows for both as the minimum values increase from 0 to 2 for both. Both standard deviations slightly increase.

Parental education is correlated with child's educational outcome. From Chapter 2 it is obvious that the effect of parental education on child's behavior can be observed on innate (e.g., IQ) as well as on acquired characteristics. The acquired characteristics are discussed in Chapter 2 where the study by Guryan *et al.* (2008) suggests that more educated parents spend more time with their children than less educated parents. This is supported by the study Kalil *et al.* (2012). The study by Neidell (2000) claims that children with parents who spend more time with them have greater human capital in terms of cognitive and non-cognitive outcomes. From that follows that children of more educated parents should have greater human capital as their parents spend more time with them. Moreover, highly educated parents are usually wealthier and can afford better schools than less educated. Concerning the sample, percentages of the highest grade completed by biological parents of the respondents are reported in Table 3.8.

Table 3.8: Biological parents' highest grade completed (%)

	Mothers	Fathers		Mothers	Fathers
2^{nd} grade	0.32	0.16	12^{th} grade	34.48	36.51
3^{rd} grade	0.54	0.64	1^{st} year college	8.40	6.37
4^{th} grade	0.54	1.02	2^{nd} year college	13.87	10.55
5^{th} grade	0.75	0.91	3^{rd} year college	8.41	2.52
6^{th} grade	2.30	2.68	4^{th} year college	13.60	13.17
7^{th} grade	0.59	1.02	5^{th} year college	3.21	1.98
8^{th} grade	2.09	2.36	6^{th} year college	4.12	4.60
9^{th} grade	3.00	3.32	7^{th} year college	0.64	1.45
10^{th} grade	4.01	3.53	8^{th} year college or more	1.23	2.94
11^{th} grade	3.91	4.28			

From Table 3.8 we can see that among both fathers and mothers the most often represented level of education is 12^{th} grade which corresponds to completing high school. 88.0% of mothers and 80.1% of fathers graduated from high school and either finished their studies or continued.

3.4 Correlation between education attainment variables

As we already mentioned, parental education is correlated with child's educational outcome. Among the respondents 84.4 % have graduated from high school as we know from Table 3.4. When we look at the gender, 87.4 % of females and 81.6 % of males graduated from high school. Those portions are comparable to those of their parents (88.0 % for mother and 80.1 % for fathers). To understand the associations between variables, we design the contingency tables. Moreover, we derive the odds ratio and phi-coefficient to interpret the strength of association between examined variables.

The odds ratio is the simplest measure of association. It presents the ratio of the odds of graduation from high school by respondents in the presence of the graduation by mother or father and the odds of graduation from high school by respondents in the absence of the graduation by mother or father. Phi-coefficient is a measure of association for two binary variables. Phi-coefficient is defined as the square root of the ratio of the chi-squared and the total of observations.

The following contingency tables display the multivariate frequency distribution of our variables. We use binary variables whether the respondents and their parents graduated from high school or not.

Table 3.9: Contingency table - respondents & mothers

		HS M		
		no	yes	Total
HS	no	113	179	292
	yes	224	1352	1579
	Total	337	1531	
Odds ratio: 3.81				
Phi coef.: 0.231				

Note: HS - graduation from high school by respondents, HS M - graduation from high school by mothers.

Table 3.10: Contingency table - respondents & fathers

		HS F		
		no	yes	Total
HS	no	121	171	292
	yes	251	1325	1579
	Total	372	1496	
Odds ratio: 3.73				
Phi coef.: 0.232				

Note: HS - graduation from high school by respondents, HS F - graduation from high school by fathers.

The cells of the tables give the counts of respondents that share each combination of high school graduation. Based on the odds ratios and phi-coefficients, we can conclude that there is a positive association in our variables, i.e., in respondents' and mothers' as well as fathers' graduation from high school.

3.5 Summary

Table 3.11 summarizes the name, type and usage of chosen variables. Dependent variable represents whether the respondent graduated from high school, explanatory variables are designing the level of parental involvement, and the control variables control the demographics.

Table 3.11: Variables: summary

	Name	Type
Dependent	HS	dummy
Explanatory	Mon M	dummy
	Mon F	dummy
	Limit Y	dummy
	Limit P	dummy
	Broke Y	dummy
	Broke P	dummy
Control	Sex	factor
	Race	factor
	Location	factor
	HH size	integer
	hgc M	integer
	hgc F	integer

Note: HS - the information about whether the respondent completed high school or not; Mon M - monitoring by mother; Mon F - monitoring by father; Lim Y - limit setting reported by youth; Lim P - limit setting reported by parent; Broke Y - Limit breaking reported by youth; Broke P - limit breaking reported by parent; HH size - household size; hgc M - highest grade completed by mother; hgc F - highest grade completed by father.

Chapter 4

Methodology

The chapter describes in detail the empirical approach. First, we specify the base model and discuss the dependent variable, a vector of explanatory variables and a vector of control variables. Second, we introduce the matching approach and discuss a vector of covariates, parameter of interest, propensity score, and subclassification method with its matching diagnostics. Finally, we discuss robust standard errors.

4.1 Model specification

The goal of our analysis is to estimate the effect of parental involvement on children's educational outcome – whether they completed high school or not. The following equation presents our baseline model:

$$y_i = \beta_0 + ParentalInvolvement_i^{\prime}\gamma + X_i^{\prime}\delta + u_i, \tag{4.1}$$

where y_i is the dependent variable corresponding to an indicator that respondent i has completed high school, and i = 1, ..., N, where N is the number of respondents. $ParentalInvolvement_i$ is a vector of variables measuring parental involvement of respondent i, X_i is a vector of control variables for respondent i, i.e., other individual or their parents' characteristics that might influence educational outcomes, and u_i is the unobserved error. A vector of variables measuring parental involvement includes monitoring by mother and father, limit setting reported by youth and parent, and limit breaking reported by youth and parent. A vector of control variables includes sex, ethnicity, household size, location (urban/rural), and the highest degree completed by both parents.

The outcome variable, child's educational outcome, is a binary variable determining whether the respondent completed high school or not. Using linear (OLS) regression is not relevant with binary outcome since several assumptions are violated (e.g., continuous outcome, normally distributed errors, homoscedasticity, ..). To model dichotomous dependent variable, logit and probit models are appropriate. They fit a nonlinear function to the data to solve the linear regression problems. The logit model uses cumulative distribution function (cdf) of the logistic distribution and the probit model uses cdf of the standard normal distribution. They both result in a similar (not identical!) output. We choose logit model for the estimation so the results will be presented in the odds ratios. Additionally, we will present marginal effects.

When regressing a binary variable determining whether the respondent graduated from high school or not on a vector of parental involvement variables, we might be facing an endogeneity problem. The endogeneity problem may occur since the variables measuring parental involvement might be correlated with the high school graduation variable as well as with the error term. Endogeneity causes the OLS coefficients to be biased. Thus, we include a vector of control variables in Equation 4.1. We include contingency tables depicting the relationship between high school graduation and categorical control variables in Appendix C. Those variables were chosen based on a study by Bogenschneider (1997). The author provides a clear summary of the literature dealing with the characteristics of the person and of the context which affect child's education attainment. Those characteristics are gender, parent's education, family structure, and ethnicity. We already control the family structure as we are dealing only with the families where both parents are present. In Chapter 2, we discussed how is the education of parents important. We also mentioned there that Zaff *et al.* (2017) associate high school dropouts with neighborhoods and ethnicity. Lastly, we described the distribution of the time on housework activities which slightly differs with a different household size. Thus, our control vector contains sex, ethnicity, size of the household, location (urban/rural), and parental level of education.

To deal with potential nonlinear relationship between control variables and the outcome variable and to be sure that we compare otherwise similar children whose parents are either highly or weakly involved, we introduce a matching method by matching similar respondents according to their characteristics. Respondents are distributed in groups by sex, ethnicity, size of household, location (urban/rural), and parental level of education. Our estimation is similar to analyzing treatment effect in medicine. The treatment variable is represented by parental involvement variables.

In general, for respondents with a high parental involvement, we find comparable respondents who have similar observable characteristics among those with a low parental involvement. We estimate the potential outcome, the potential average effect of parental involvement on the high school graduation of the respondents. This idea is described in detail in the following subchapters.

4.2 Matching Approach

In matching method terminology, we are interested in the causal effect of the treatment (T=1), in our case "high parental involvement", relative to no treatment (T=0), in our case "low parental involvement", on the child's educational outcome. For parental involvement we use separately 6 treatment variables which are equal to the components of the previously mentioned vector of explanatory variables, i.e., $ParentalInvolvement_i$. It includes monitoring child by mother, monitoring child by father, limit setting answered by youth as well as by parent, and limit breaking also answered by youth and by parent.

Paper by Stuart (2010) merges broad and diverse literature dealing with matching methods, and it focuses on original work on matching methods as well as on new ideas across disciplines. The paper defines the matching method as a method, whose goal is to balance the covariates' distribution among treated and control groups.

In our case, the vector of covariates is equal to the vector of control variables which includes sex, ethnicity, household size, location (urban/rural), and the highest degree completed by parents. The treated group contains those individuals whose parents are very involved, and the control group consists of individuals with a low parental involvement. There is no possibility to observe outcome for a single respondent with both treatment and no treatment as single respondent is either treated (parents are very involved) or not treated (parents are not very involved). That is why we have to equate as much as possible the distribution of covariates among treated and control groups. Rosenbaum & Rubin (1983) called this estimation of the causal effect as a missing data problem, since always one value for each individual is missing, either treatment or no treatment. This unobserved outcome is called counterfactual outcome.

Propensity score matching estimation of the effect of the treatment can be done in two key steps – the first one is design, and the second one is outcome analysis. The first step is a key tool for matching methods. It uses only the information about the respondents, in our case sex, ethnicity, household size, location (urban/rural), and the highest degree completed by parents. Here, in the first step, the outcome variable is not used at all. After completion of the first step, the second step begins – comparison of the outcomes from the treated and control groups. In the following subchapters we discuss the methodology in more detail.

4.2.1 Potential outcome approach

As discussed in Caliendo & Kopeinig (2005), the main focus should be on individuals and their characteristics – vector of covariates, treatment (parental involvement), and outcome (completed high school or not). As we discussed before, respondents with high parental involvement are indicated as treated (T=1), and control (T=0) otherwise. The outcomes are then $Y_i(T_i)$ for each individual i, where i = 1, ..., N and N is number of respondents. Then the treatment effect can be understood as:

$$\tau_i = Y_i(1) - Y_i(0) \tag{4.2}$$

In Equation 4.2, τ_i represents individual causal effect, the treatment effect for an individual i. It should be noted that $Y_i(T_i)$ is non-random as in non-experimental studies are individuals placed into the groups by many factors which may affect also the outcome. As we discussed before, for each individual i only one of the potential outcomes can be observed. Here occurs the fundamental evaluation problem. Since we cannot observe both outcomes for each individual, we are interested in average treatment effects (or population average treatment effects). To estimate average treatment effects, the treatment effect for individual i has to be independent of others treatment participation.

4.2.2 Parameter of interest

Before introducing the matching method, we should choose the primary treatment effect of interest. The most common are Average Treatment Effect on the Treated (ATT), which is defined as the average effect of the treatment for those who receive the treatment, and Average Treatment Effect (ATE), which is defined as the average treatment effect in the population from which we have a random sample. They are called estimands.

The estimands control how the subclasses are created and how the weights are computed. If estimand is set to ATT, the subclassification is based on quantiles of the distance measure in the treated group. On the other hand, if estimand is set to ATE, subclassification is based on quantiles of the distance measure in the full sample. For our further analysis we use ATT. ATT is defined as:

$$\tau_{ATT} = E(\tau|T = 1) = E[Y(1)|T = 1] - E[Y(0)|T = 1] \tag{4.3}$$

As we mentioned before, the problem arises here. $E[Y(0)|T = 1]$ represents the unobserved - counterfactual - mean for those being treated. To complete an estimation of ATT, we have to substitute the counterfactual mean for those being treated. As we already know, individuals are selected into treatment groups by many factors that may or may not influence the outcome. That is why using $E[Y(0)|T = 0]$ (mean outcome of untreated) is not appropriate as the factors are most likely to affect the treatment decision as well as the outcome variable we are interested in. In other words, the outcomes from both groups would differ even if we did not introduce the treatment variable, which would lead to a self-selection bias. By regrouping Equation 4.3 and adding $-E[Y(0)|T = 0]$ to both sides of the equation we obtain:

$$E[Y(1)|T = 1] - E[Y(0)|T = 0] = \tau_{ATT} + E[Y(0)|T = 1] - E[Y(0)|T = 0], \tag{4.4}$$

where $E[Y(0)|T = 1]-E[Y(0)|T = 0]$ represents the self-selection bias. The parameter τ_{ATT} is defined only if $E[Y(0)|T = 1] - E[Y(0)|T = 0] = 0$, meaning that the self-selection bias does not occur. To solve the problem in Equation 4.4, we introduce propensity score matching further in this chapter. Rosenbaum & Rubin (1983) claim, speaking of observational (non-experimental) studies, that adjustment for the propensity score is an important component of the analysis as the evidence of residual bias in the propensity score means evidence of potential bias in the estimated treatment effects. Then Rosenbaum & Rubin (1983) introduce three standard techniques for adjustment. One of them is subclassification which is the most suitable for our purposes and we will describe it further.

4.2.3 Propensity score

The propensity score ($P(\mathbf{X}) = P(T = 1|\mathbf{X})$) is the probability of being treated for an individual given the covariates $\mathbf{X}$. The propensity score is one of the balancing scores. According to Rosenbaum & Rubin (1983), if potential outcomes are independent of the treatment conditional on the covariates $\mathbf{X}$, then they are also independent of the treatment conditional on a b($\mathbf{X}$), the balancing score.

Assumptions for estimating strategy:

1. **Positivity (Common Support):** assignment is probabilistic:

$$0 < P[T_i = 1|\mathbf{X}, Y(1), Y(0)] < 1,$$

or

$$0 < P[T_i = 1|\mathbf{X}] < 1.$$

2. **No unmeasured confounding:**

$$P[T_i|\mathbf{X}, Y(1), Y(0)] = P[T_i|\mathbf{X}],$$

further - unconfoundedness given propensity score:

$$P[T_i|P(\mathbf{X})].$$

First assumption is also known as overlap condition. Let $\mathbf{X}$ be a set of observable covariates which are not affected by the treatment (i.e., sex, ethnicity, household size, location, and highest degree completed by parents). Common support ensures that individuals with the same covariates' values (i.e., the same $\mathbf{X}$ values) have positive probability of being both - treated and non treated. If P($\mathbf{X}$)=0 or P($\mathbf{X}$)=1 for some $\mathbf{X}$, then the individuals with such $\mathbf{X}$ are either never treated or always treated and we cannot use matching. Second assumption, also called conditional independence assumption or unconfoundedness, assumes that potential outcomes are independent of treatment conditional on vector of covariates $\mathbf{X}$, or further on propensity score P($\mathbf{X}$).

If both assumptions hold, the propensity score matching estimator for ATT is simply the mean difference of outcomes weighted by respective propensity scores. When estimating ATT, sufficient assumptions are $P[T_i = 1|\mathbf{X}] < 1$, and $P[T_i|\mathbf{X}, Y(0)] = P[T_i|P(\mathbf{X})]$. τ_{ATT}^{PSM} is then mathematically defined as [5]:

$$\tau_{ATT}^{PSM} = E_{P(\mathbf{X})|T=1}\{E[Y(1)|T = 1, P(\mathbf{X})] - E[Y(0)|T = 0, P(\mathbf{X})]\}. \quad (4.5)$$

Estimating the Propensity score

Two important choices have to be made to estimate the propensity score. The estimation model and the variable in such a model. Let start with description of the variables as we already mentioned them previously.

According to Caliendo & Kopeinig (2005), the model of propensity score should include only variables which are not affected by the treatment, and the covariates $\mathbf{X}$ have to satisfy assumption (2) (outcomes are independent of treatment conditional on the propensity score). Also omitting important variables leads to bias in the resulting estimates. In other words, we should include variables which affect both the treatment decision and the outcome variable. The best way how to prevent violating these assumptions is to choose variables which are fixed over the time or which are measured before the treatment. Our set of covariates meets these conditions. Since we do not assume that sex, ethnicity, and the highest degree completed by parents are likely to change. Concerning household size and location, those covariates could change. However, we measure them at the same point of time as the treatment variables. There are several ways how to evaluate whether the selection of variables is appropriate. Two most common are Hit or Miss Method and Statistical Significance.

With regard to model choice, we can use any discrete choice model. Speaking of binary treatment, then probit and logit models are preferred and they usually derive similar results. In our case, we use logit model for estimating the propensity score.

4.2.4 Subclassification

The propensity score subclassification is performed using the `MatchIt` package in R, which is the only package used for the propensity score estimation.[6]

[5] Propensity Score Matching (PSM)

[6] For a more detailed description of MatchIt package in R see: `https://cran.r-project.org/web/packages/MatchIt/MatchIt.pdf`

We use the subclassification matching method which places units from treatment and control groups into subclasses based on quantiles of the propensity score. Unlike other matching methods, subclassification uses all individuals from the treatment as well as from the control group. There are no discarded control individuals, all of the observations are used in the analysis. When estimand is set to ATT, then are the quantiles based on the propensity score distribution in the treated group. Computed weights are then the portions of treated units respective to the total in each subclass (control units plus treated units within the subclass).

Generally, the estimation of marginal effects after subclassification is either computed for each subclass separately and then their average is calculated or we can observe marginal effect by using respective subclass weights in full sample, which is also known as Marginal Mean Weighting through Stratification (MMWS). Both methods should yield the same (or similar) result. Regarding the binary outcomes, as in our case, the estimation of the marginal effects slightly differ. Binary outcomes are measured by the risk difference, risk ratio, or odds ratio. For better interpretation and understanding the results, we introduce the odds ratio (OR) by exponentiating the coefficients of logit model. The OR is a noncollapsible effect measure, so the computation of marginal effect estimate is done by computing the average of the predicted subclass-specific odds under each treatment from which is then computed the marginal effect estimate.

Orihara & Hamada (2021) discuss the optimal number of subclasses for subclassification on the propensity score. They select the number of subclasses which minimizes Mean Squared Error (MSE) of the subclassification estimator. According to Rosenbaum & Rubin (1983), 90% of bias is removed by only five subclasses based on the propensity score. However, this is not general conclusion. Five subclasses are not universal recommendation and may not always be optimal. Even though it is not always optimal, in our case it is suitable. We choose five subclasses based on matches diagnostics which are described in the following subchapter.

Balance Assessment

Diagnosing the quality of matched samples is the most important step when using the matching methods. When matching is done, it should be followed by evaluation of the covariate balance whether are groups matched correctly.

Austin *et al.* (2005) and Austin *et al.* (2007) have proposed as the most appropriate statistical method for assessing balance the standardized mean differences for balance assessment in observational studies. The Standardized Mean Difference (SMD) is the difference in means of each covariate between treated and control groups standardized by a standardization factor. When targeting ATT, the standardization factor is the standard deviation of the covariate in the treated group. So the standardized difference is defined as:

$$d = \frac{100 \times |\overline{x}_T - \overline{x}_C|}{\sqrt{\frac{s_T^2 + s_C^2}{2}}}, \tag{4.6}$$

where $\overline{x}_T$ and $\overline{x}_C$ are the means of the variables among the treated and control subjects, respectively. s_T^2 and s_C^2 are the sample standard deviantions of covariate in the treated and control subjects, respectively. Noah Greifer, the author of the vignette for `MatchIt` package, recommended other possible measures to asses balance including variance ratios, empirical cumulative density functions (eCDF) statistics, or visual diagnostics.[7]

In Table 4.1 we can see the standardized mean difference for each covariate prior to matching and for matched sample. The value of standardized mean difference close to zero indicates good balance. Vignettes for `MatchIt` package recommend tresholds 0.1 and 0.05 for prognostically important covariates. Furthermore, in Appendix D we can find means of each covariate for treated as well as for control group before and after matching. We can also find there graphical diagnostics (love plots of standardized mean differences and histograms of propensity scores). As the SMDs of all covariates after the matching are below 0.1, even below 0.05 except the SMD of highest grade completed by mother for limit setting variable responded by parent, we can conclude that the balance of the means between treated and control group significantly improved after the matching. This is supported by the table in Appendix D.

[7]The vignette described in detail can be found here: `https://cran.r-project.org/web/packages/MatchIt/vignettes/assessing-balance.html`

Table 4.1: Standardized mean differences

	SMD BEFORE SUBCLASSES					
	Mon M	Mon F	Lim Y	Lim P	Br Y	Br P
distance	0.4601	0.3605	0.3208	0.3107	0.2523	0.3884
sexM	-0.1369	0.0631	0.0835	0.0565	0.1794	0.3575
sexF	0.1369	-0.0631	-0.0835	-0.0565	-0.1794	-0.3575
raceblack	-0.2354	-0.2168	0.1327	0.2349	0.0229	0.0952
racehispanic	-0.1352	-0.0996	0.0743	0.0590	0.0953	-0.0470
racemixed	-0.0370	-0.0059	-0.0617	0.0165	-0.0108	-0.0420
racenonBH	0.2911	0.2387	-0.1598	-0.2367	-0.0963	-0.0340
HH size	-0.0940	-0.0302	0.1414	0.0934	0.0970	0.0239
urban0	0.1105	0.1820	0.0668	0.0182	-0.0937	-0.0806
urban1	-0.0624	-0.1647	-0.0849	-0.0262	0.0830	0.0814
urban2	-0.1066	-0.0282	0.0459	0.0199	0.0130	-0.0110
hgc M	0.2609	0.1370	-0.2067	-0.1215	-0.1058	0.0181
hgc F	0.2696	0.2266	-0.2059	-0.1348	-0.0988	-0.0273
	SMD AFTER SUBCLASSES					
	Mon M	Mon F	Lim Y	Lim P	Br Y	Br P
distance	0.0426	0.0168	0.0163	0.0155	0.0279	0.0312
sexM	-0.0191	0.0000	-0.0009	-0.0078	0.0057	0.0210
sexF	0.0191	-0.0000	0.0009	0.0078	-0.0057	-0.0210
raceblack	-0.0168	-0.0274	0.0004	0.0440	0.0049	0.0198
racehispanic	-0.0128	0.0016	0.0157	-0.0400	0.0232	-0.0009
racemixed	0.0093	0.0074	-0.0160	-0.0140	-0.0075	-0.0114
racenonBH	0.0216	0.0172	-0.0115	0.0001	-0.0222	-0.0138
HH size	0.0209	0.0046	0.0170	-0.0055	0.0267	-0.0023
urban0	-0.0006	-0.0196	-0.0020	0.0091	-0.0028	-0.0286
urban1	-0.0056	0.0171	0.0005	0.0044	0.0043	0.0272
urban2	0.0149	0.0046	0.0030	-0.0297	-0.0036	0.0000
hgc M	0.0322	0.0002	-0.0178	0.0506	-0.0244	0.0038
hgc F	0.0207	0.0208	-0.0049	0.0248	-0.0258	0.0092

Note: horizontal: Mon M - monitoring by mother; Mon F - monitoring by father; Lim Y - limit setting reported by youth; Lim P - limit setting reported by parent; Broke Y - limit breaking reported by youth; Broke P - limit breaking reported by parent. Vertical: sexM - male; sexF - female; racenonBH - Non-Black/Non-Hispanic; HH size - household size; urban0 - rural; urban1 - urban;, urban2 - unknown; hgc M - highest grade completed by mother; hgc F - highest grade completed by father.

4.2.5 Robust and Cluster-Robust Standard Errors

Usage of the cluster-robust standard errors is preferred in the most cases of the matching method. However, concerning the subclassification, it is the opposite. The cluster-robust standard errors are appropriate when a large number of clusters is present, which subclassification method does not have. In the case of subclass method, regular robust standard errors are appropriate when estimating marginal effects because of a few clusters. What is more, the robust standard errors are necessary when using weights for the estimation of the treatment effect. Model based standard errors assume the weights to be frequency weights and not the probability weights. Thus, it is inappropriate to use them and we have to implement the robust standard errors.[8]

[8]For more details about estimating effects after matching and(cluster-) robust standard errors see vignettes for MatchIt package: `https://cran.r-project.org/web/packages/MatchIt/vignettes/estimating-effects.html`

Chapter 5

Results

In this chapter we discuss the results. Firstly, we derive sample description. Secondly, we introduce results from a simple logistic regression. Thirdly, we implement subclassification matching method which is the main scope of this thesis. Finally, we compare the results and provide discussion.

The log odds (i.e., the coefficients from a logistic regression) lack a meaningful metric and present useful information solely about the sign of the effect. Since the exact size of the effect of the log odds is difficult to imagine, we introduce the odds ratios and the marginal effects for the interpretation of the results. The outputs (log odds) from a logistic regressions are included in the Appendix E.

5.1 Sample description

The sample consists of 1,868 participants, of whom 1,093 is monitored by mother, 971 is monitored by father, 874 youths feel like their parents are setting them limits, 818 of parents reported that they are setting limits to youth, 812 youths reported that they have broken the limit, and 557 parents think that their child did break the limit.

From a simple data observing we can see that mothers monitor their children more than fathers. The limit setting variable reported by respondents is about 3% higher than the same variable reported by parents. In addition, more respondents reported limit breaking than their parents.

Regarding our sample, females are more likely to graduate from high school. About 10.7 % of Black women and nearly 42 % of Black men did not complete high school. Concerning Hispanic race, the portions are 20.5 % and 33.6 % for women and men, respectively. Nearly 13.8 % of White women and 16.5 % of White men did not graduate from high school. Speaking of mixed race, the portions are not good indicators as there are only 10 respondents of a mixed race (100 % of women and 50 % of men of mixed race did complete high school).

The data suggests that the indicator for completing high school tends to differ with respect to respondents' observable characteristics. As other have shown, respondents who graduated from high school tend to have more educated parents and are more likely living in a household with less persons. In Table 5.1 we can see means with SDs in parentheses for respondents who have finished high school and those who have not. We can notice that the means of parental education are higher for those respondents who have completed high school.

Table 5.1: Mean and SD by high school status

	HS	
	yes	no
HH size	4.69	4.90
	(1.26)	(1.48)
hgc M	13.17	11.40
	(2.91)	(2.91)
hgc F	13.17	11.11
	(3.20)	(3.12)

Note: HH size - household size (range: <2;14>); hgc M - highest degree completed by mothers (range: <2;20>); hgc F - highest degree completed by father (range: <2;20>).

In addition, we provide three tables in Appendix F. First depicts the means for parental involvement variables before the dummy modification. The values are presented for males and females separately. Besides, we can see the means also separately for those who have completed high school and those who have not. In the second and third table we present means with standard deviations in parentheses for variables of household size and the highest degree completed by both parents. We compare the means for those who are treated (i.e., high parental involvement) and those who are not (i.e., low parental involvement). Concerning the variable of monitoring, we can see that more educated parents tend to monitor their children more. Different result is observed only for the highest degree completed by mother when we look at father's monitoring variable for female respondents. What is surprising is that from limit setting variable follows that less educated parents are setting the limits more. Respondents who broke the limit tend to have less educated parents. The only exception is observed for a variable of the highest degree completed by father for both male and female respondents. Speaking of the household size, there is no observable pattern.

5.2 Simple Logistic Regression

First, the results of a simple logistic regression are presented, without matching. It estimates the effects of parenting styles as we keep other child and parental characteristics constant, without caring for the common support. The logistic regressions are performed separately for each parental involvement variable. The dependent variable is a binary educational variable determining whether the respondent graduated from high school or not. We choose as a treshold for statistical significance a p-value less than 0.05 (i.e., 5 % level of significance).

Odds ratios

In this section, we present the odds ratios and 95 % confidence interval (CI). To obtain those results, we exponentiate the log odds as well as the confidence intervals. The 95 % confidence interval for the ln(OR) is computed as: 95% CI for $ln(OR) = ln(OR) + 1.96 \times \{SEln(OR)\}$, the 95 % confidence interval for the OR is then computed by exponentiating the results from 95% CI for ln(OR).

The odds ratios and the confidence intervals are reported in Table 5.2. The vector of control variables was included in the regression, however, we present only the coefficients concerning the variables of parental involvement. The CIs tell us whether the results are statistically significant or not. The CI for the variable of limit setting reported by youths includes the value of 1 meaning that the result is not statistically significant. The CIs for monitoring youth by father and limit setting reported by parents are close to the value of 1.

The values of the exponentiated coefficients for parental monitoring are greater than 1, meaning that the odds of graduating from high school increase when respondent is monitored by mother or father. From Table 5.2 we can see that the effect is greater when youth is monitored by mother than by father. The odds of completing high school for youths are higher by a factor of 1.61 and 1.38 when monitored by mother and father, respectively. In other words, there is a 38-61% increase in the odds of graduation from high school when monitored by parents. Concerning the limit breaking variables, there is a 39-42% decrease in the odds of graduating from high school when respondent breaks the limit. What is surprising is the sign of the odds for limit setting variables. Setting the limits to the respondents by their parents decreases the odds of graduation from high school. However, only the limit setting variable responded by parents is statistically significant (at the given significance level). Nevertheless, the opposite effect can be caused by that more problematic children have probably more limits from their parents than those who are not problematic. Moreover, the results might be caused by the limitation of our sample since we regard only to Americans born in 1983 and 1984.

Table 5.2: Logit model - OR

	OR	2.5%	97.5%
mon_M1	1.61	1.20	2.15
mon_F1	1.38	1.06	1.80
lim_Y1	0.89	0.69	1.16
lim_P1	0.68	0.48	0.95
broke_Y1	0.61	0.47	0.79
broke_P1	0.58	0.44	0.76

Note: mon_M - monitoring by mother; mon_F - monitoring by father; lim_Y - limit setting reported by youth; lim_P - limit setting reported by parent; broke_Y - limit breaking reported by youth; broke_P - limit breaking reported by parent, and 1 represents that low parental involvement is a base level.

Marginal Effects

For further interpretation of the coefficients, we introduce the marginal effects. The marginal effect of categorical explanatory variables corresponds to an effect of a one-unit discrete change in the category in such a variable. For our variables of parental involvement, the base level refers to "low parental involvement". The marginal effect of statistically significant coefficients ranges from -0.07 to 0.06. The results show that the expected probability of completing high school is higher for those who are monitored by parents (i.e., treated) than for those who are not monitored. The probabilities are higher by 4 and 6 percentage points for monitoring by mothers and father variables, respectively, all other explanatory variables held constant. Limit breaking decreases the probability of completing high school by 6 to 7 percentage points.

Concerning the variables included in the control vector, we can see that the variables of gender, race, and the highest grade completed by parents provide statistically significant results. The variables for household size and neighborhood (i.e., location - urban/rural) show no effect at any level of significance. The expected probability of graduating from high school is higher for females than males by 5 to 6 percentage points.

Table 5.3: Logit model - marginal effects

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
mon_M1	0.06^{**}					
	(0.02)					
mon_F1		0.04^{*}				
		(0.02)				
lim_Y1			-0.01			
			(0.02)			
lim_P1				-0.04^{*}		
				(0.02)		
broke_Y1					-0.06^{***}	
					(0.02)	
broke_P1						-0.07^{***}
						(0.02)
sexF	0.06^{***}	0.06^{***}	0.06^{***}	0.06^{***}	0.06^{***}	0.05^{**}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
racehispanic	0.05^{**}	0.06^{**}	0.06^{**}	0.06^{**}	0.06^{**}	0.05^{**}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
racemixed	-0.05	-0.06	-0.06	-0.06	-0.06	-0.07
	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)
racenonBH	0.02	0.02	0.03	0.02	0.02	0.02
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
HH_size	-0.00	-0.00	0.00	0.00	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
urban1	0.00	0.00	-0.00	-0.00	0.00	0.00
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
urban2	-0.02	-0.02	-0.03	-0.03	-0.02	-0.02
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
hgc_F	0.02^{***}	0.02^{***}	0.02^{***}	0.02^{***}	0.02^{***}	0.02^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
hgc_M	0.01^{***}	0.01^{***}	0.01^{***}	0.01^{***}	0.01^{***}	0.01^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Num. obs.	1868	1868	1868	1868	1868	1868
Log Likelihood	-738.30	-740.55	-743.02	-740.74	-736.46	-735.77
Deviance	1476.59	1481.10	1486.03	1481.48	1472.92	1471.54
AIC	1498.59	1503.10	1508.03	1503.48	1494.92	1493.54
BIC	1559.45	1563.96	1568.89	1564.33	1555.78	1554.40

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Coefficient (marginal effect dy/dx) for categorical variable is the discrete change from the base.

Robust standard errors are provided in parentheses.

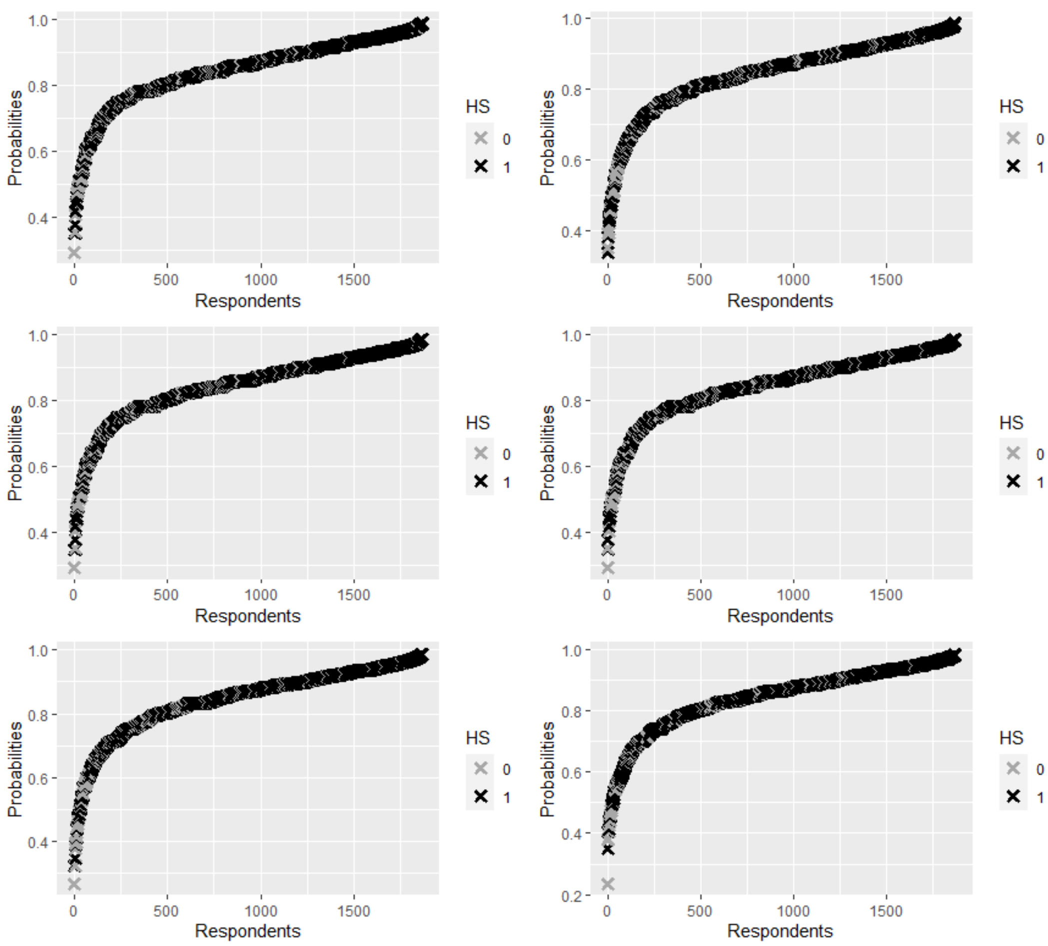
Note: mon_M - monitoring by mother; mon_F - monitoring by father; lim_Y - limit setting reported by youth; lim_P - limit setting reported by parent; broke_Y - limit breaking reported by youth; broke_P - limit breaking reported by parent, and 1 represents that low parental involvement is a base level; sexF - female; racehispanic - Hispanic respondents; racemixed - mixed race respondents; racenonBH - Non-Black/Non-Hispanic; HH_size - household size; urban1 - urban; urban2 - unknown; hgc_F - highest grade completed by father; hgc_M - highest grade completed by mother.

5.2.1 Predicted probabilities

Figure 5.1 shows the predicted probabilities based on a logistic regression models for completing high school given each parental involvement variable. In the first row, we can observe the predicted probabilities for completing high school based on logistic models using monitoring by mothers and fathers variables, from the left to the right, respectively. In the second and third row are depicted the predicted probabilities for completing high school for limit setting and limit breaking variables, respectively. In the first column are the probabilities for the variables responded by youth, in the second column are the probabilities for the variables responded by parents.

We can see that most of the respondents who graduate from high school are predicted to have a high probability of completing high school, and most of the respondents who do not graduate from high school are predicted to have a low probability of completing high school.

Figure 5.1: Predicted probabilities for completing high school



5.3 Matching Approach

5.3.1 Estimating the Propensity Score

The baseline characteristics of the treated and the control participants are sex, ethnicity, household size, location (urban/rural), and the highest degree completed by parents. None of the covariates have standardized mean differences that exceeds the value of 0.1. Moreover, only one slightly exceeds the value of 0.05. The systematic differences in baseline characteristics between the treated and the control respondents are reduced or eliminated by matching.

For the estimation of the propensity score we use logit model and five subclasses. In Table 5.4 we can see the number of respondents in each group, either treated or control group, for each of the treatment variable. Further, in Table 5.5 we can observe the number of respondents in each subclass and respective subclasses' weights.

Table 5.4: Number of control vs treated

	mon_M	mon_F	lim_Y	lim_P	broke_Y	broke_P
Control	408	897	994	453	1056	1311
Treated	1460	971	874	1415	812	557

Note: mon_M - monitoring by mother; mon_F - monitoring by father; lim_Y - limit setting reported by youth; lim_P - limit setting reported by parent; broke_Y - limit breaking reported by youth; broke_P - limit breaking reported by parent.

Table 5.5: Weights & number of respondents

mon_M		mon_F		lim_Y		lim_P		broke_Y		broke_P	
Weight	N	Weight	N	Weight	N	Weight	N	Weight	N	Weight	N
0.565	144	0.635	281	0.620	321	0.702	129	0.721	285	0.666	396
0.995	82	0.919	196	0.994	199	0.749	121	0.888	243	0.734	356
1	1460	0.979	184	1	874	1	1415	1	812	1	557
1.063	77	1	971	1.048	190	1.147	79	1.009	210	1.215	215
1.329	61	1.210	139	1.214	164	1.224	74	1.262	167	1.233	208
1.867	44	1.962	97	1.659	120	1.812	50	1.404	151	1.973	136

Note: mon_M - monitoring by mother; mon_F - monitoring by father; lim_Y - limit setting reported by youth; lim_P - limit setting reported by parent; broke_Y - limit breaking reported by youth; broke_P - limit breaking reported by parent.

Initial propensity score models are estimated by using the vector of covariates. To estimate the propensity scores, the logistic regression models are used in which treatment statuses, i.e., parental involvement dummies, are regressed on the baseline characteristics, i.e., vector of covariates. This is performed independently for each variable of parental involvement, i.e., we obtain six propensity score models.

The estimated propensity scores are then the predicted probabilities of exposure to the treatment (parental involvement) from the logistic regression models. Once the propensity scores have been estimated for each subject, treated and control subjects are matched on the respective propensity scores.

5.3.2 Subclassification

We present the odds ratios and the marginal effects for the interpretation of the results. The log odds are included in Appendix E.

Odds ratios

In Table 5.6 are reported the odds ratios and the CIs. We can see that the CIs for the variables of limit setting reported by both youths and parents include the value of 1 meaning that the results are not statistically significant. Speaking of the variable of monitoring by father, the lower bound is close to 1. The odds of graduating from high school when the respondent is monitored by mother is 1.68 greater than for the respondent who is not monitored by mother. Meaning that there is a 68% increase in the odds of graduating from high school when monitored by mother. As for fathers, the increase in the odds is by 32%. Speaking of the limit breaking variables, there is a 39-41% decrease in the odds of graduation from high school when respondent breaks the limit.

Table 5.6: Subclassification - OR

	OR	2.5%	97.5%
mon_M1	1.68	1.26	2.23
mon_F1	1.32	1.02	1.71
lim_Y1	0.91	0.71	1.16
lim_P1	0.74	0.54	1.00
broke_Y1	0.61	0.48	0.78
broke_P1	0.59	0.46	0.77

Note: mon_M - monitoring by mother; mon_F - monitoring by father; lim_Y - limit setting reported by youth; lim_P - limit setting reported by parent; broke_Y - limit breaking reported by youth; broke_P - limit breaking reported by parent, and 1 represents that low parental involvement is a base level.

In Table 5.7 we can see the marginal treatment effects with the robust standard errors in parentheses. As we discussed in Chapter 4, the robust standard errors are necessary when estimating the treatment effect.

The base level refers to "low parental involvement" for our variables of parental involvement. All of the observed coefficients are statistically significant and they range from -0.08 to 0.10. The treatment represented by monitoring variables increases the expected probability of completing high school. The variable of monitoring youth by mother shows the largest marginal effect of 0.10. Meaning that the expected probability of youth completing high school is greater by 10 percentage points for those who are monitored by mother. When monitored by father, the expected probability of youth completing high school is greater only by 6 percentage points. Breaking limits decreases the expected probability of youth completing high school by 8 percentage points for both observed variables.

Marginal Effects

Table 5.7: Subclassification - marginal effects

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
mon_M1	0.10^{***}					
	(0.02)					
mon_F1		0.06^{**}				
		(0.02)				
lim_Y1			-0.04^{*}			
			(0.02)			
lim_P1				-0.05^{**}		
				(0.02)		
broke_Y1					-0.08^{***}	
					(0.02)	
broke_P1						-0.08^{***}
						(0.02)
Num. obs.	1868	1868	1868	1868	1868	1868
Log Likelihood	-798.40	-804.39	-807.62	-805.66	-799.15	-800.36
Deviance	1596.80	1608.78	1615.24	1611.32	1598.30	1600.72
AIC	1600.80	1612.78	1619.24	1615.32	1602.30	1604.72
BIC	1611.86	1623.85	1630.30	1626.39	1613.36	1615.78

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Robust standard errors are provided in parentheses.

Note: mon_M - monitoring by mother; mon_F - monitoring by father; lim_Y - limit setting reported by youth; lim_P - limit setting reported by parent; broke_Y - limit breaking reported by youth; broke_P - limit breaking reported by parent, and 1 represents that low parental involvement is a base level.

5.4 Comparison

The odds ratios and the confidence intervals for the logistic regression as well as for the matching method are reported in Table 5.8. We observe the largest difference in the effects of the variable describing monitoring by mother. We can conclude that the variable for monitoring the youth by mother has greater effect (by 7 %) when estimated by matching method than by logistic regression. Moreover, the variable is more statistically significant. Concerning the variable for monitoring youth by father, both estimations are statistically significant. The result is lowered by 6 % when estimated by matching method. Both estimations are lower than the estimations for mothers. Speaking of limit setting variables, only the CI for limit setting by parent in a logistic regression do not involve the value of 1. Variables for breaking the limits reported by parents and respondents are comparable from both regressions.

Table 5.8: Comparison - OR

		OR	2.5%	97.5%
mon_M1	Logit	1.61	1.20	2.15
	Matching	1.68	1.26	2.23
mon_F1	Logit	1.38	1.06	1.80
	Matching	1.32	1.02	1.71
lim_Y1	Logit	0.89	0.69	1.16
	Matching	0.91	0.71	1.16
lim_P1	Logit	0.68	0.48	0.95
	Matching	0.74	0.54	1.00
broke_Y1	Logit	0.61	0.47	0.79
	Matching	0.61	0.48	0.78
broke_P1	Logit	0.58	0.44	0.76
	Matching	0.59	0.46	0.77

Note: mon_M - monitoring by mother; mon_F - monitoring by father; lim_Y - limit setting reported by youth; lim_P - limit setting reported by parent; broke_Y - limit breaking reported by youth; broke_P - limit breaking reported by parent, and 1 represents that low parental involvement is a base level.

In Table 5.9 we can see the marginal effects from the logistic regression as well as the marginal effects from the subclassification method. In the table we report the variables included in the vector of parental involvement. We can see that the results are more statistically significant when estimated by matching method. Moreover, the marginal effects are greater with matching than the marginal effects from a simple logistic regression.

Table 5.9: Comparison - marginal effects

		Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
mon_M1	Logit	0.06^{**} (0.02)					
	Matching	0.10^{***} (0.02)					
mon_F1	Logit		0.04^{*} (0.02)				
	Matching		0.06^{**} (0.02)				
lim_Y1	Logit			-0.04 (0.02)			
	Matching			-0.04^{*} (0.13)			
lim_P1	Logit				-0.04^{*} (0.02)		
	Matching				-0.05^{**} (0.02)		
broke_Y1	Logit					-0.06^{***} (0.02)	
	Matching					-0.08^{***} (0.02)	
broke_P1	Logit						-0.07^{***} (0.02)
	Matching						-0.08^{***} (0.02)

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Robust standard errors are provided in parentheses.

Note: mon_M - monitoring by mother; mon_F - monitoring by father; lim_Y - limit setting reported by youth; lim_P - limit setting reported by parent; broke_Y - limit breaking reported by youth; broke_P - limit breaking reported by parent, and 1 represents that low parental involvement is a base level.

The surprising signs of the marginal effects of the limit setting variables hold in both cases. However, when reported by youth, the significant effect is observed only by matching. Nevertheless, these findings might be caused by the limitations of our data. We have to keep in mind that we have limited sample. We focus on a sample of American youth born in 1983 and 1984

5.5 Discussion

The problem in the estimation of the effect of parental involvement on the child's educational outcome is that there might be an issue that more educated parents use more appropriate parenting techniques. They might monitor their children in a more "friendly way" which is more comfortable for the child. They may set the limits more accordingly to the child. Thus, children of more educated parent might have less intentions to break the limits.

In the end we do not know whether these are the techniques that influence children's education outcomes or the educational outcomes of more educated parents are better because children inherited higher IQ from their parents. Parental education is correlated with child's educational outcome as we already discussed. The effect of parental education on child's behavior can be observed on innate (e.g., IQ) as well as on acquired characteristics. Moreover, Guryan *et al.* (2008) and Kalil *et al.* (2012) suggest that more educated parents spend more time with their children than less educated parents. The study by Neidell (2000) claims that children with parents who spend more time with them have greater human capital in terms of cognitive and non-cognitive outcomes. From that follows that children of more educated parents should have greater human capital as their parents spend more time with them. That may result in different education outcome.

When we control for those innate and acquired characteristics, we should obtain more precise results. The matching approach is a tool how to observe the pure effect of parental involvement since we control the baseline characteristics. We control the innate characteristics (e.g., IQ) by the covariates of the level of parental education. Assuming that the vector of covariates explains parental choice of parental involvement, matching approach will help identify the treatment effect. According to the results of matching method, the parental involvement do affect the child's educational outcome. Nevertheless, the results might be limited since we have restricted sample of Americans born in 1983 and 1984.

The surprising results of the limit setting variables might be caused not only by the limitations of our data. There is a chance that more problematic children have more limits. This "problematicity" is hidden in the error term but it might affect the probability of graduation from high school and at the same time it might affect limit setting variable. It may be a source of endogeneity. We suggest to include the variables of child's criminal behavior for further research.

Chapter 6

Conclusion

The main goal of this thesis was to investigate the effect of parental involvement on children's educational outcome, specifically, whether they completed high school or not. To observe such an effect, we used a sample from the National Longitudinal Survey of Youth 1997. The National Longitudinal Survey of Youth 1997 is a publicly open database. Our sample consisted of American respondents born in 1983 and 1984 who lived with both birth parents in a household when they were from 12 to 14 years old. We investigated three main issues. Firstly, whether parental involvement in youth's life affects their graduation from high school. Secondly, we compared the effects of maternal and paternal involvement on child's educational outcome. Finally, we researched on the impact of limit breaking by respondents on their educational attainment.

The research results were conducted by a simple logistic regression and the matching method. The matching method - subclassification on propensity score - controls for the observable child's and parents' characteristics which can affect the outcome, i.e., the completion of high school, as well as the parental involvement variables.

We can conclude that there are observable significant effects in child's educational outcome caused by parental involvement in child's life. The monitoring variables depict the increase in the expected probability of youth completing high school. Using the subclassification matching method, the probabilities are greater by 10 and 6 percentage points for mothers' and fathers' monitoring, respectively. Buchanan *et al.* (1992) claim that degree of monitoring has impact on child's scholastic achievement. We support this statement.

The research on the impact of limit breaking by respondents on their educational attainment produces statistically significant effects. Based on the subclassification matching method, limit breaking by youth decreases the probability of completing high school by 8 percentage points.

The results of our analysis provides mostly the effects we expected. The exception is that maternal involvement has a greater impact than paternal. Moreover, we found surprising results for limit setting variables. The limit setting variables decrease the odds of completion of high school. We though that the sign would be otherwise. This might be caused by the limitations of our data. We have to keep in mind that we have limited sample. We focus on a sample of American youth born in 1983 and 1984.

6.1 Suggestions for future research

The limit breaking variable could be significant because of the discrepancies. About 36 % of youths from the original (i.e., not reduced) dataset claim that they did break at least one of the limits even though their parent reported that they did not break any of these three limits. According to Erford (1995) and Eccles *et al.* (1991), observing parallel questions is very useful as discrepancies across the answers indicate misunderstanding in who in fact sets the limits, which often leads to limits breaking by youths from parents' point of view. This thesis did not treat uniquely those discrepancies. However, NLSY97 offers more of parallel questions, not only about limit setting and limit breaking. Another parallel questions which were asked youth as well as parents are: parents' marital relationship, concerning whether mother/father is supportive of father/mother, and behavioral and emotional problems scale for both girls and boys, where scale indicates frequency of behavioral problems of the youth.

There are numerous studies arguing that parents' marital relationship can affect children's outcome. As regards to behavior problems, they are linked to family and neighborhood characteristics. Moreover, we can predict youth behavior problems based on parents' outcomes - academic achievement and employment. Further research can investigate the relationship whether the parental supportiveness has positive impact on children's educational outcome. In addition, we can examine whether the discrepancies in the answers of the parallel questions negatively affect children's education outcome.

Another issue which deserves more attention is whether higher level of limit setting leads to higher level of limit breaking since "forbidden fruit is the sweetest". Also, for the observation of such an effect, it could be appropriate to include the variables for child's "problematicity".

Moreover, in Chapter 2 we discussed that the limits are giving to children the feeling of parental care. Children with a few limits or even completely without the rules experience anxiety. In further research, we can focus whether there is a correlation between limit setting, anxiety, and child's educational outcome.

Bibliography

Austin, P. C., P. Grootendorst, & G. M. Anderson (2007): "A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study." *Statistics in Medicine* **26(4)**: pp. 734–753.

Austin, P. C., M. M. Mamdani, T. A. Stukel, G. M. Anderson, & J. V. Tu (2005): "The use of the propensity score for estimating treatment effects: administrative versus clinical data." *Statistics in Medicine* **24(10)**: pp. 1563–1578.

Belfield, C. R. & H. M. Levin (2007): "The price we pay: economic and social consequences of inadequate education." *Washington, DC: The Brookings Institution* .

Bogenschneider, K. (1997): "Parental Involvement in Adolescent Schooling: A Proximal Process with Transcontextual Validity." *Journal of Marriage and Family* **59(3)**: pp. 718–733.

Bornstein, M. H. (2005): *Handbook of Parenting: Volume 4 Social Conditions and Applied Parenting.* Taylor & Francis.

Bryan, J. (2005): "Fostering educational resilience and achievement in urban schools through school-family-community partnerships." *Professional School Counseling* **8**: p. 219.

Buchanan, C. M., E. E. Maccoby, & S. M. Dornbusch (1992): "Adolescents and their families after divorce: Three residential arrangements compared." *Journal of Research on Adolescence* **2(3)**: pp. 261–291.

Caliendo, M. & S. Kopeinig (2005): "Some Practical Guidance for the Implementation of Propensity Score Matching." *Discussion Paper No. 1588* .

COLE, S. A. (2017): *The Impact of Parental Involvement on Academic Achievement*. Ph.D. thesis, Northcentral University.

ECCLES, J. S., C. M. BUCHANAN, C. FLANAGAN, A. FULIGNI, C. MIDGLEY, & D. YEE (1991): "Control Versus Autonomy During Early Adolescence." *Journal of Social Issues* **47(4)**: pp. 53–68.

ERFORD, B. T. (1995): "Parental Autonomy-Enhancement and Development of Self-Efficacy." *Psychological Reports* **77(3)**: pp. 1347–1353.

FEINSTEIN, L. & J. SYMONS (1999): "Attainment in Secondary School." *Oxford Economic Papers* .

GORDON, K. (2017): "4 Types of Limits That Children Need." *Hand in Hand* .

GURYAN, J., E. HURST, & M. KEARNEY (2008): "Parental Education and Parental Time with Children." *Journal of Economic Perspectives* **22(3)**: pp. 23–46.

HOOVER-DEMPSEY, K. V. & H. M. SANDLER (1997): "Why Do Parents Become Involved in TheirChildren's Education? ." *Vanderbilt University* **67(1)**: pp. 3–42.

KALIL, A., R. RYAN, & M. COREY (2012): "Diverging Destinies: Maternal Education and the Developmental Gradient in Time with Children." *Demography* **49(4)**: pp. 1361–1383.

MCCALLUMORE, K. M. & E. F. SPARAPANI (2010): "The importance of the ninth grade on high school graduation rates and tudent succes in high school." *Education* **130(3)**.

MCDOWELL, K., A. JACK, & M. COMPTON (2018): "Parent Involvement in Pre-Kindergarten and the Effects on Student Achievement." *The Advocate* **23**.

MCLANAHAN, S. (2004): "Diverging destinies: How children are faring under the second demographic transition." *Demography* **41**: pp. 607–627.

MORIN, A. (2020): "How to Discipline Kids With Positive and Negative Consequences." *Verywell Family* .

Morin, A. (2021): "Reasons Why It Is Important to Set Limits With Kids." *Verywell Family* .

Neidell, M. J. (2000): "Early Parental Time Investments In Children's Human Capital Development: Effects Of Time In The First Year On Cognitive And Non-Cognitive Outcomes." *UCLA Economics Working Papers 806, UCLA Department of Economics* .

Orihara, S. & E. Hamada (2021): "Determination of the optimal number of strata for propensity score subclassification." *Statistics & Probability Letters* **168**: p. 108951.

Rasmussen, A. W. (2009): "Allocation of Parental Time and the Long-Term Effect on Children's Education." *Department of Economics and CIM, Aarhus University.* **Working paper 09-22**.

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

Snow, C. (2006): "What Counts as Literacy in Early Childhood? In: McCartney K, Phillips D, editors. Handbook of Early Child Development." *Malden, MA: Blackwell Publishing* pp. 274–294.

Stuart, E. A. (2010): "Matching methods for causal inference: A review and a look forward." *Statistical science : a review journal of the Institute of Mathematical Statistics* **25**: pp. 1–21.

Sum, A., I. Khatiwada, & J. McLaughlin (2009): "The consequences of dropping out of high school: joblessness and jailing for high school dropouts and the high cost for taxpayers." *Boston: Center for Labor Market Studies Publications* **(23)**.

Zaff, J. F., A. Donlan, A. Gunning, S. E. Anderson, E. McDermott, & M. Sedaca (2017): "Factors that Promote High School Graduation: a Review of the Literature." *Educ Psychol Rev* **29**: pp. 447–476.

Appendix A

Description of parental involvement variables

Table A.1: Descriptive statistics of parenting techniques

	Monitoring (M)	Monitoring (F)	Limit setting (Y)	Limit setting (P)	Limit breaking (Y)	Limit breaking (P)
Description	how much does mother know about kids free time - friends, friends' parents, teachers, where are he/she	how much does father know about kids free time - friends, friends' parents, teachers, where are he/she	autonomy, control & limit setting - responses by youths	autonomy, control & limit setting - responses by parents	how many times have youth broken the limits in past 30 days - responses by youths	how many times have youth broken the limits in past 30 days - responses by parents
Values	$\{1,\ldots, 16\}$	$\{0,\ldots, 16\}$	$\{1,\ldots, 6\}$	$\{1,\ldots, 6\}$	$\{0, 1\}$	$\{0, 1\}$
Mean	10.75	8.51	3.41	4.33	0.43	0.30
SD	3.08	3.99	1.44	1.25	0.50	0.46
Median	11.0	9.0	3.0	4.0	0.0	0.0
Min	1	0	1	1	0	0
Max	16	16	6	6	1	1

Appendix B

Histograms

Figure B.1: Histogram

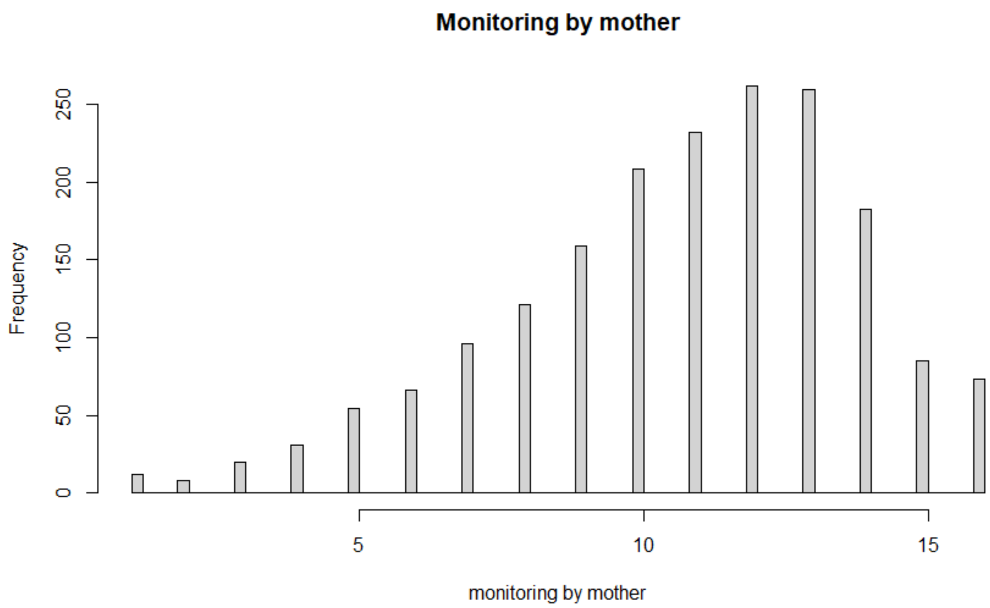


Figure B.2: Histogram

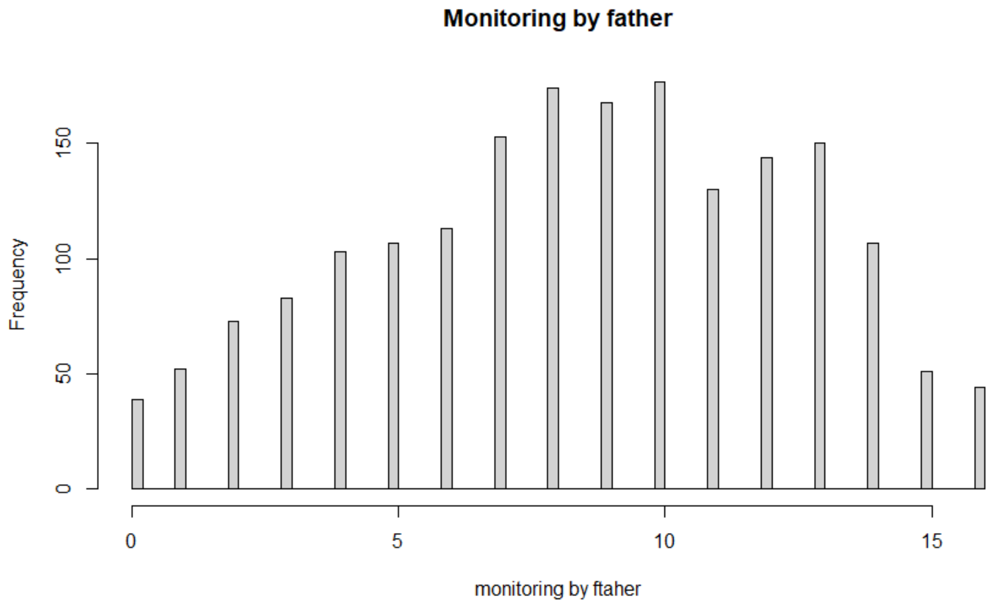


Figure B.3: Histogram

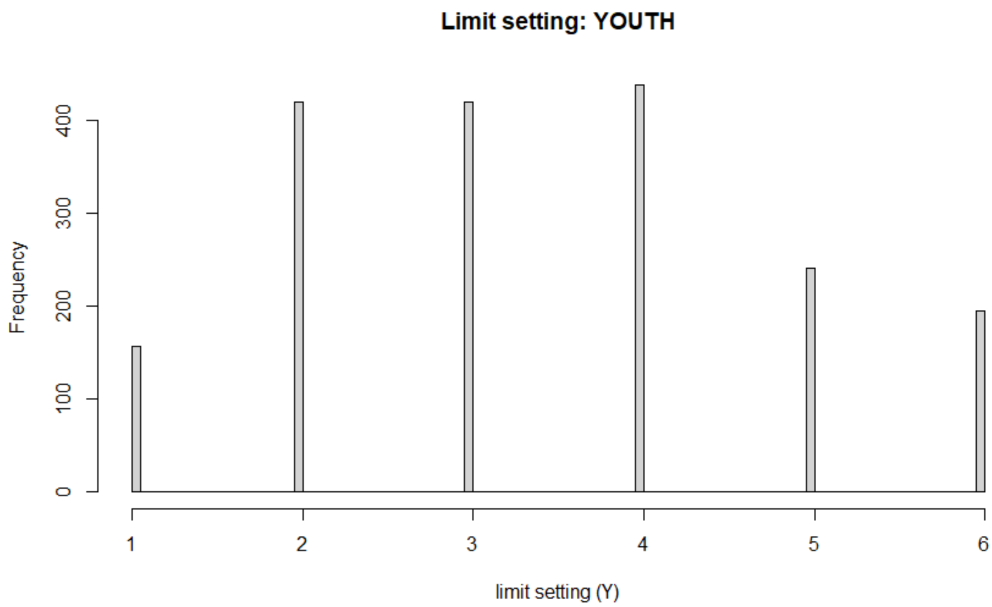


Figure B.4: Histogram

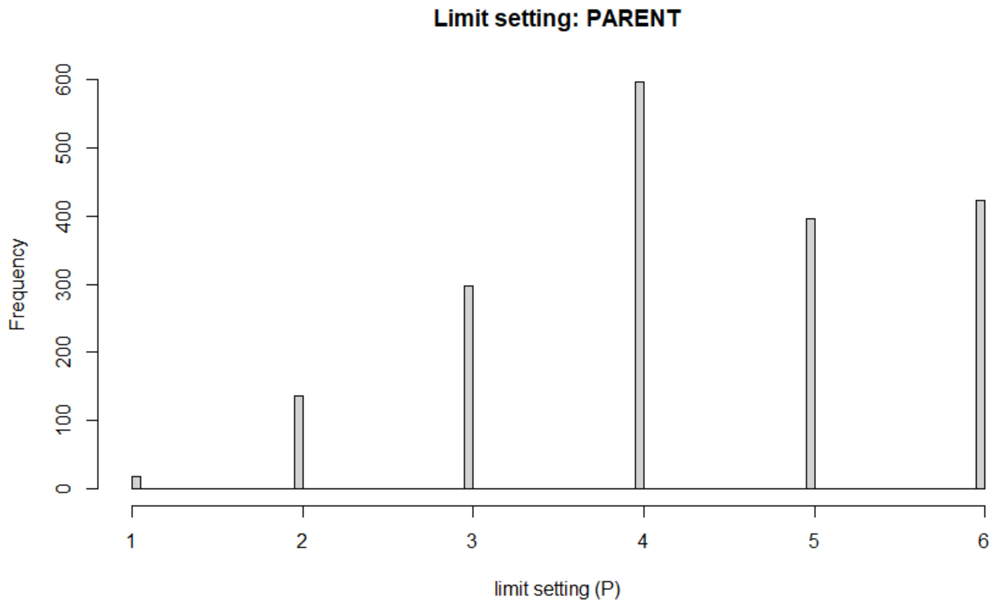


Figure B.5: Histogram

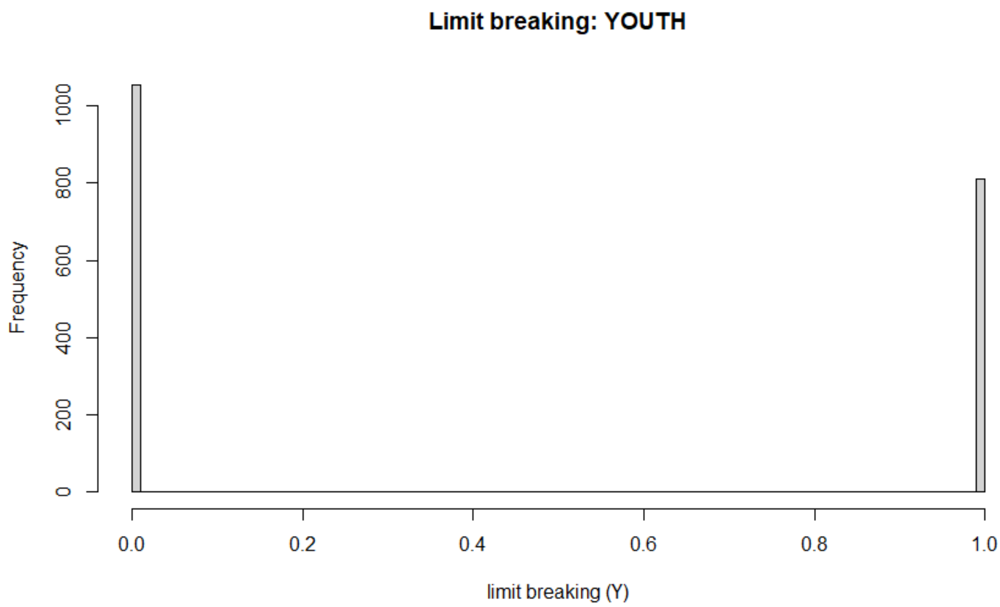
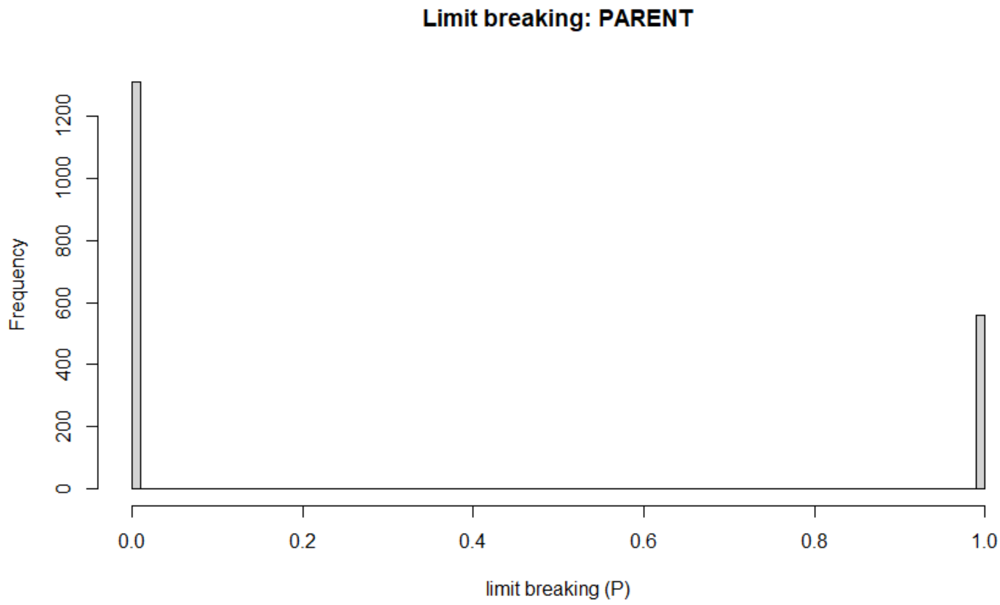


Figure B.6: Histogram



Appendix C

Contingency tables

Table C.1: Contingency table - HS & gender

		GENDER		
		no	male	female
HS	no	180	112	292
	yes	796	780	1576
	Total	976	892	
Odds ratio: 1.57				
Phi coef.: 0.081				

Note: HS - graduation from high school by respondents..

Table C.2: Contingency table - HS & ethnicity

		RACE				
		Black	Hispanic	mixed	White	
HS	no	55	76	2	159	292
	yes	242	283	8	1043	1576
	Total	297	359	10	1202	
			X^2	df	$P(> X^2)$	
Likelihood Ratio			15.100	3	0.0017329	
Pearson			15.637	3	0.0013459	

Note: HS - graduation from high school by respondents..

Table C.3: Contingency table - HS & location

		RACE			
		Rural	Urban	Unknown	
HS	no	76	200	16	292
	yes	422	1090	64	1576
	Total	498	1290	80	
		X^2	df	$P(> X^2)$	
Likelihood Ratio		1.1489	2	0.56301	
Pearson		1.2254	2	0.54189	

Note: HS - graduation from high school by respondents..

Appendix D

Balance Assessment

Table D.1: Means: treated vs. control group, before vs. after subclassification

	BEFORE SUBCLASSES											
	Mon M		Mon F		Lim Y		Lim P		Br Y		Br P	
	T	C	T	C	T	C	T	C	T	C	T	C
distance	0.79	76	0.53	0.50	0.48	0.46	0.76	0.74	0.44	0.43	0.32	0.29
sexM	0.51	0.58	0.54	0.51	0.54	0.50	0.53	0.50	0.57	0.48	0.64	0.47
sexF	0.49	0.42	0.46	0.49	0.46	0.50	0.47	0.50	0.43	0.52	0.36	0.53
raceblack	0.14	0.22	0.12	0.20	0.19	0.13	0.18	0.09	0.16	0.16	0.19	0.15
racehispanic	0.18	0.23	0.17	0.21	0.21	0.18	0.20	0.17	0.21	0.18	0.18	0.20
racemixed	0.00	0.01	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.01
racenonBH	0.67	0.54	0.70	0.59	0.60	0.68	0.62	0.73	0.62	0.66	0.63	0.65
HH size	4.69	4.81	4.70	4.74	4.82	4.63	4.75	4.62	4.80	4.66	4.74	4.71
urban0	0.28	0.23	0.31	0.22	0.28	0.25	0.27	0.26	0.24	0.28	0.24	0.28
urban1	0.68	0.71	0.65	0.73	0.67	0.71	0.69	0.70	0.71	0.67	0.72	0.68
urban2	0.04	0.06	0.04	0.05	0.05	0.04	0.04	0.04	0.04	0.04	0.04	0.04
hgc F	13.04	12.17	13.21	12.46	12.50	13.16	12.74	13.18	12.66	13.00	12.79	12.88
hgc M	13.06	12.29	13.09	12.68	12.57	13.18	12.81	13.16	12.71	13.04	12.93	12.88
	AFTER SUBCLASSES											
	Mon M		Mon F		Lim Y		Lim P		Br Y		Br P	
	T	C	T	C	T	C	T	C	T	C	T	C
distance	0.79	0.79	0.53	0.53	0.48	0.48	0.76	0.75	0.44	0.44	0.32	0.32
sexM	0.51	0.52	0.54	0.54	0.54	0.55	0.53	0.53	0.57	0.57	0.64	0.63
sexF	0.49	0.48	0.46	0.46	0.46	0.45	0.47	0.47	0.43	0.43	0.36	0.37
raceblack	0.14	0.15	0.12	0.13	0.19	0.19	0.18	0.16	0.16	0.16	0.19	0.18
racehispanic	0.18	0.19	0.17	0.17	0.21	0.20	0.20	0.21	0.21	0.21	0.18	0.18
racemixed	0.00	0.00	0.01	0.01	0.00	0.00	0.01	0.01	0.01	0.01	0.00	0.00
racenonBH	0.67	0.66	0.70	0.69	0.60	0.61	0.62	0.62	0.62	0.63	0.65	0.64
HH size	4.69	4.67	4.70	4.70	4.82	4.80	4.75	4.76	4.80	4.76	4.74	4.74
urban0	0.28	0.28	0.31	0.32	0.28	0.28	0.27	0.26	0.24	0.25	0.24	0.25
urban1	0.68	0.69	0.65	0.64	0.67	0.67	0.69	0.69	0.71	0.71	0.72	0.70
urban2	0.04	0.04	0.04	0.04	0.05	0.05	0.04	0.05	0.04	0.05	0.04	0.04
hgc F	13.04	12.97	13.2	13.14	12.50	12.52	12.74	12.66	12.66	12.75	12.79	12.76
hgc M	13.06	12.97	13.09	13.09	12.57	12.62	12.81	12.66	12.71	12.78	12.93	12.92

Note: T - treated group, C - control group, sexM - male; sexF - female; racenonBH - Non-Black/Non-Hispanic; HH size - household size; urban0 - rural; urban1 - urban;, urban2 - unknown; hgc F - highest grade completed by father; hgc M - highest grade completed by mother.

Figure D.1: Love plot - Monitoring by mother

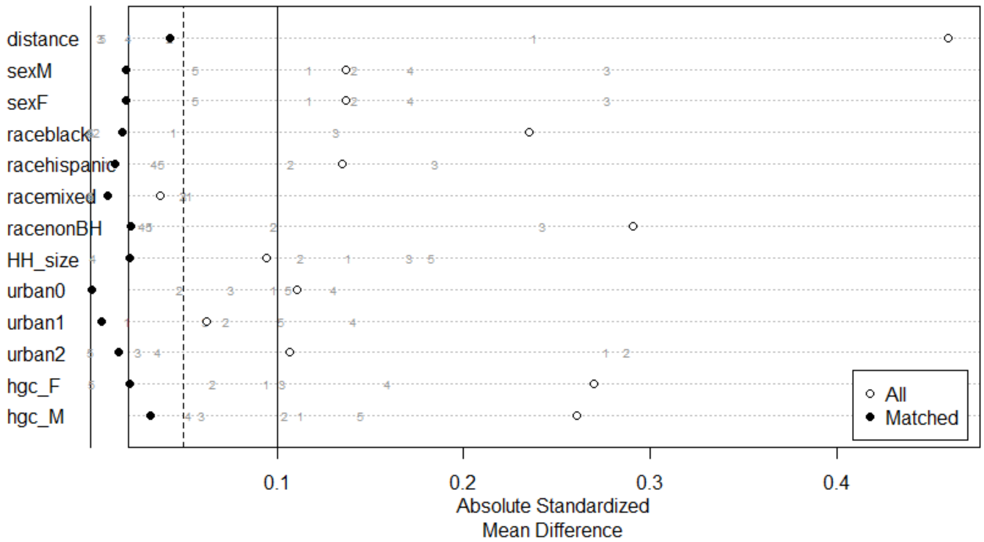


Figure D.2: Monitoring by mother

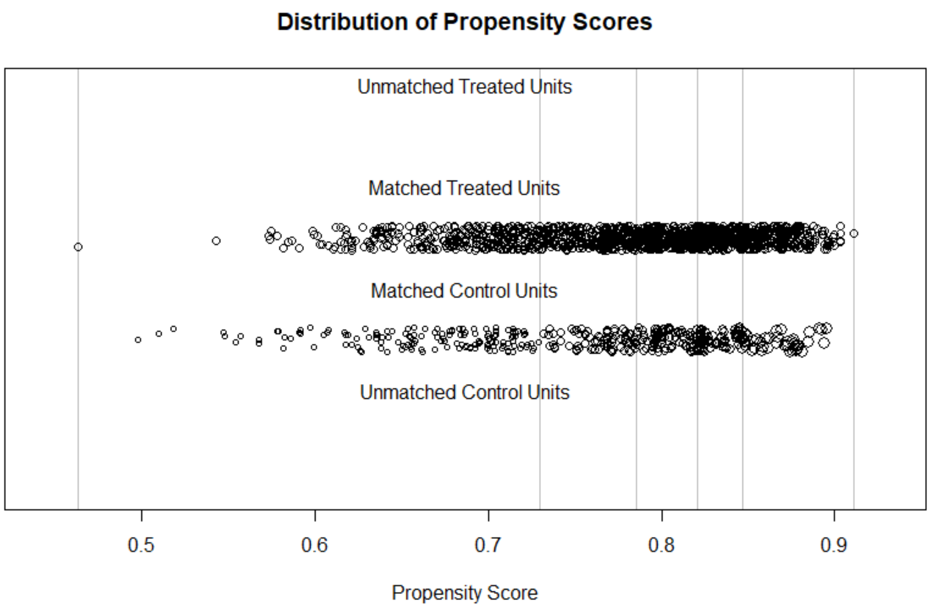


Figure D.3: Propensity Score Histograms - Monitoring by mother

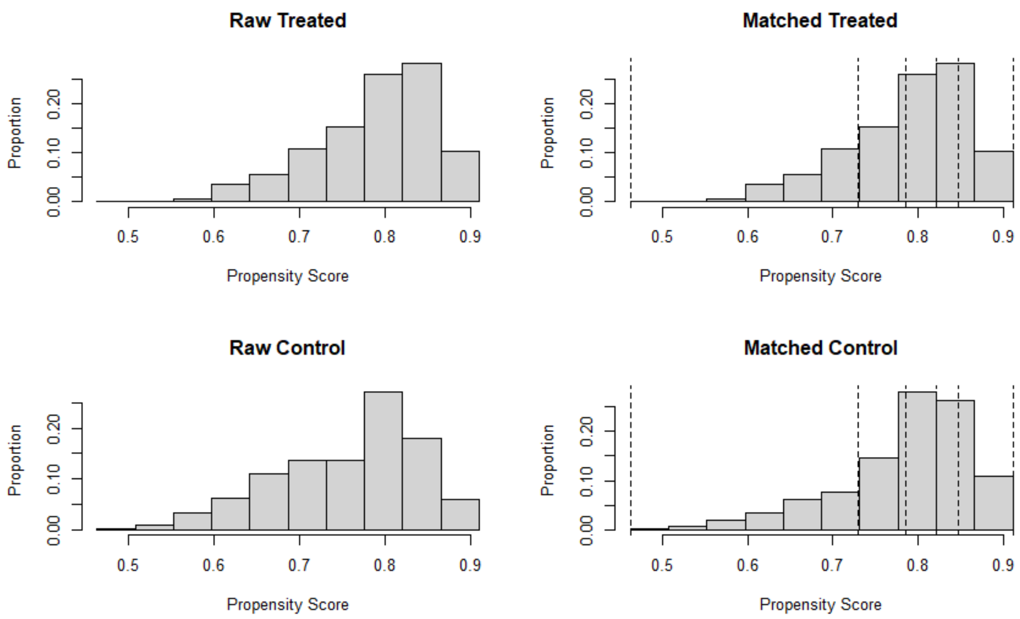


Figure D.4: Love plot - Monitoring by father

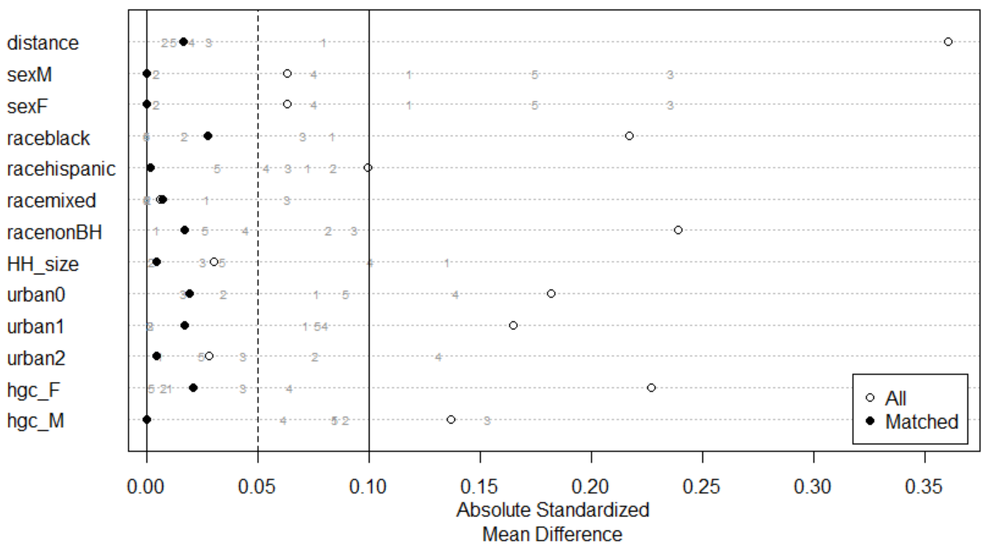


Figure D.5: Monitoring by father

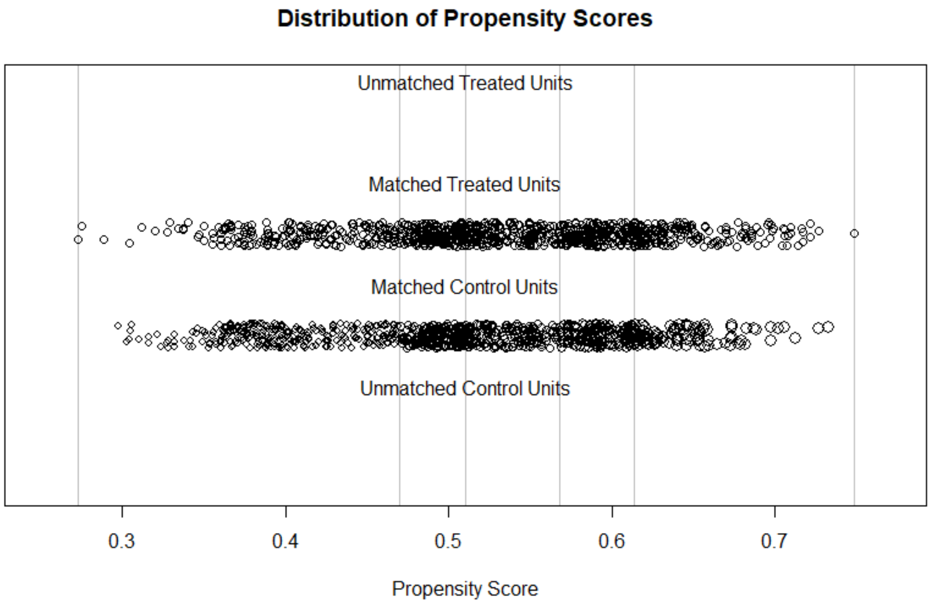


Figure D.6: Propensity Score Histograms - Monitoring by father

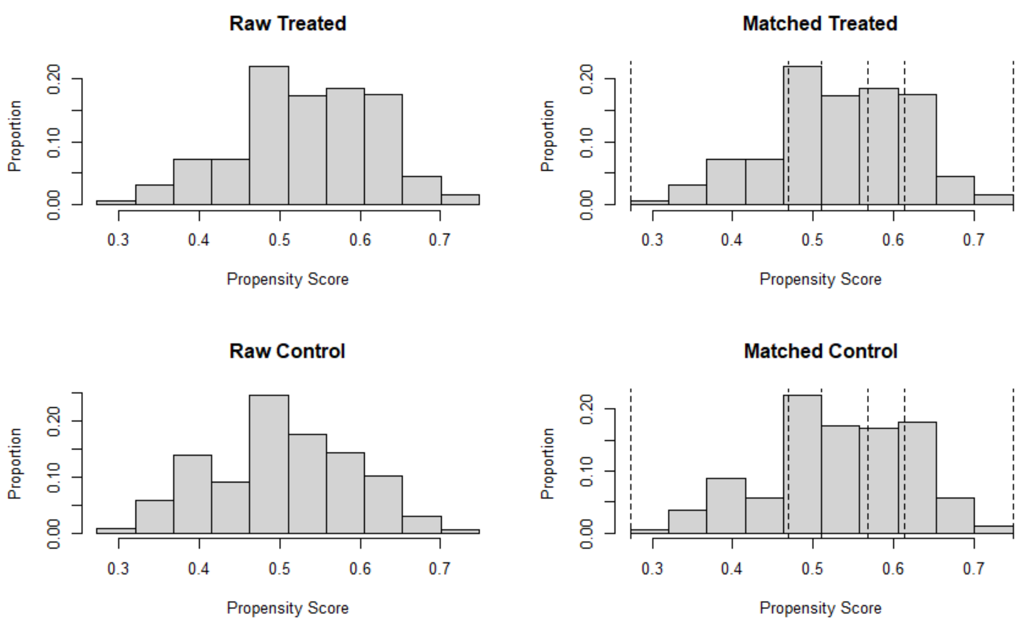


Figure D.7: Love plot - Limit setting (youth)

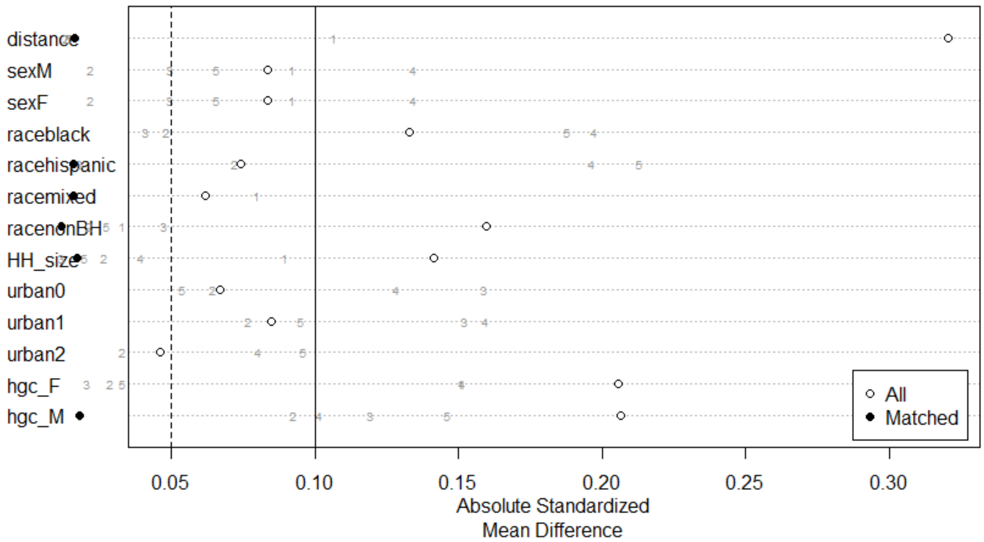


Figure D.8: Limit setting (youth)

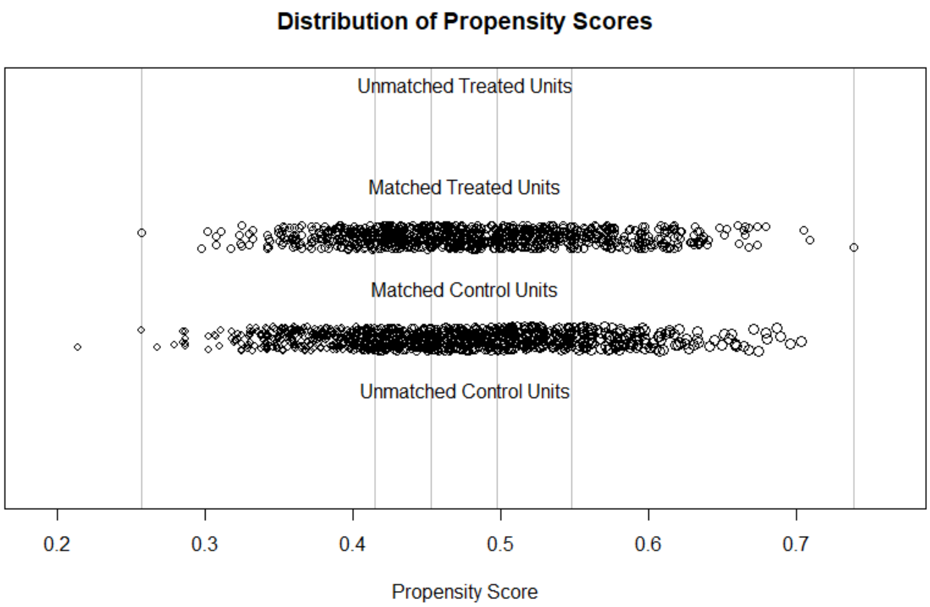


Figure D.9: Propensity Score Histograms - Limit setting (youth)

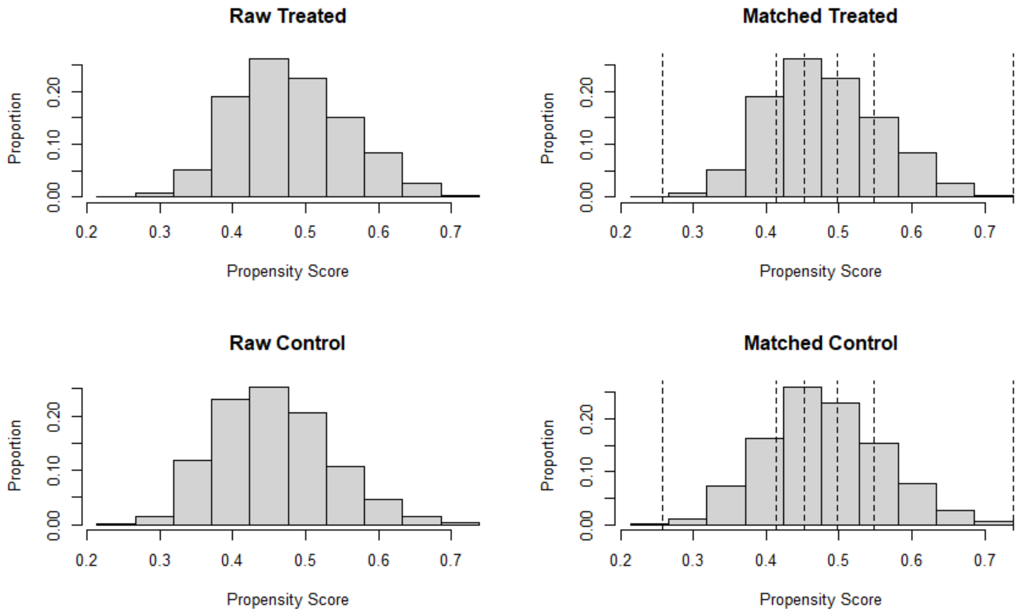


Figure D.10: Love plot - Limit setting (parent)

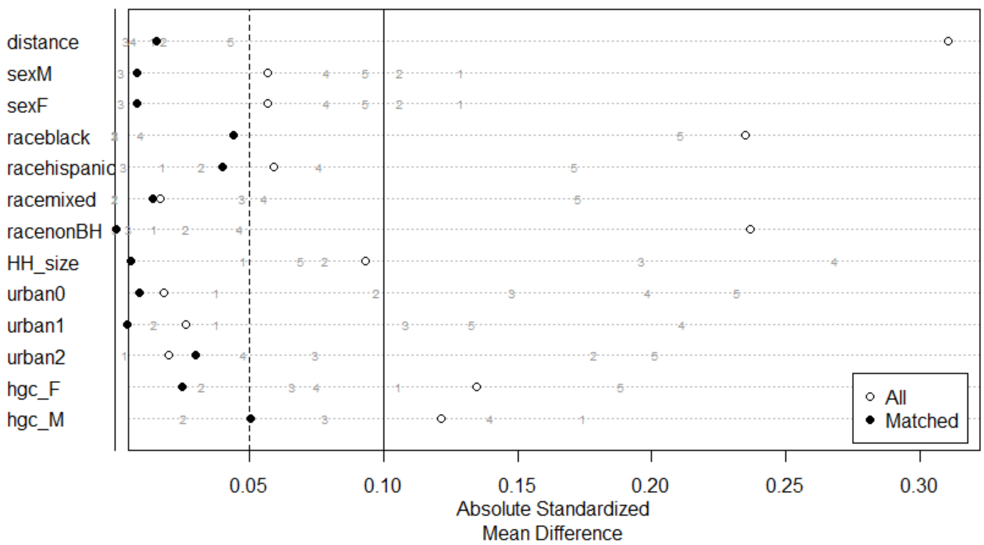


Figure D.11: Limit setting (parent)

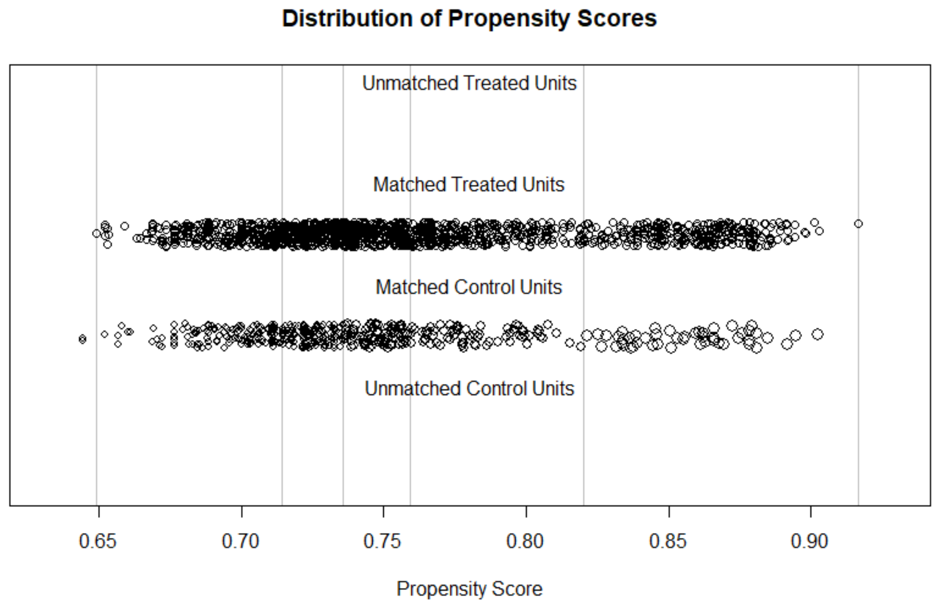


Figure D.12: Propensity Score Histograms - Limit setting (parent)

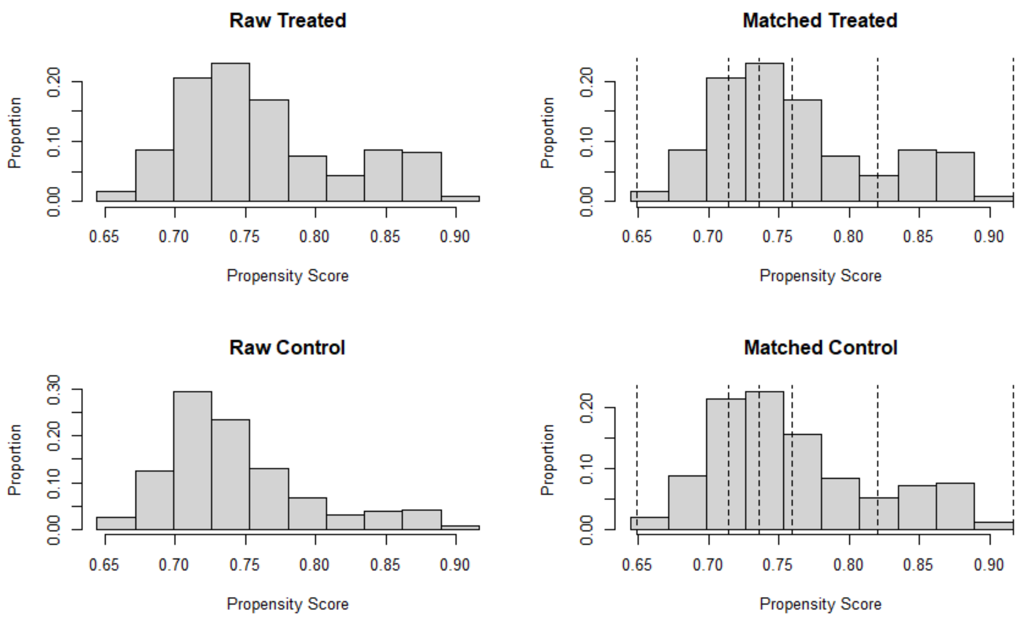


Figure D.13: Love plot - Limit breaking (youth)

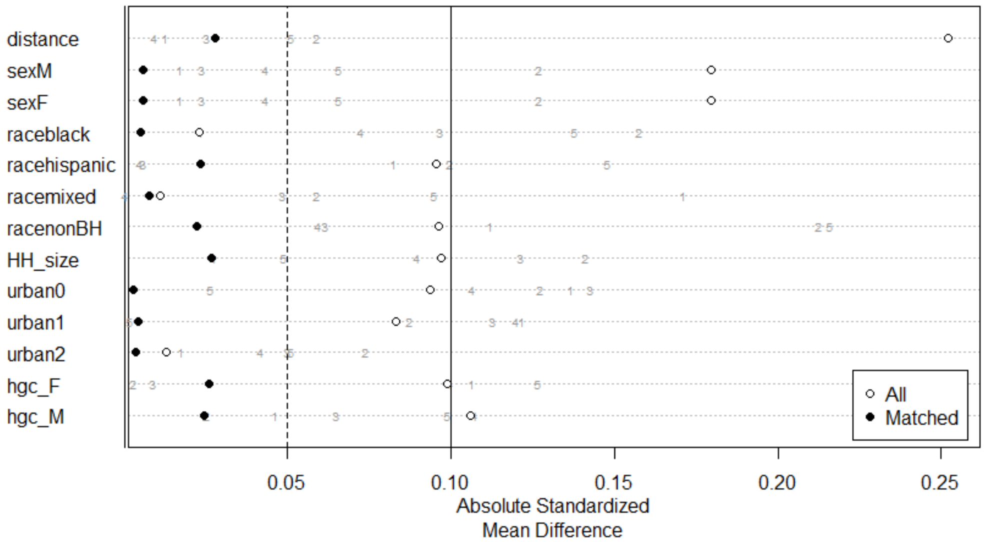


Figure D.14: Limit breaking (youth)

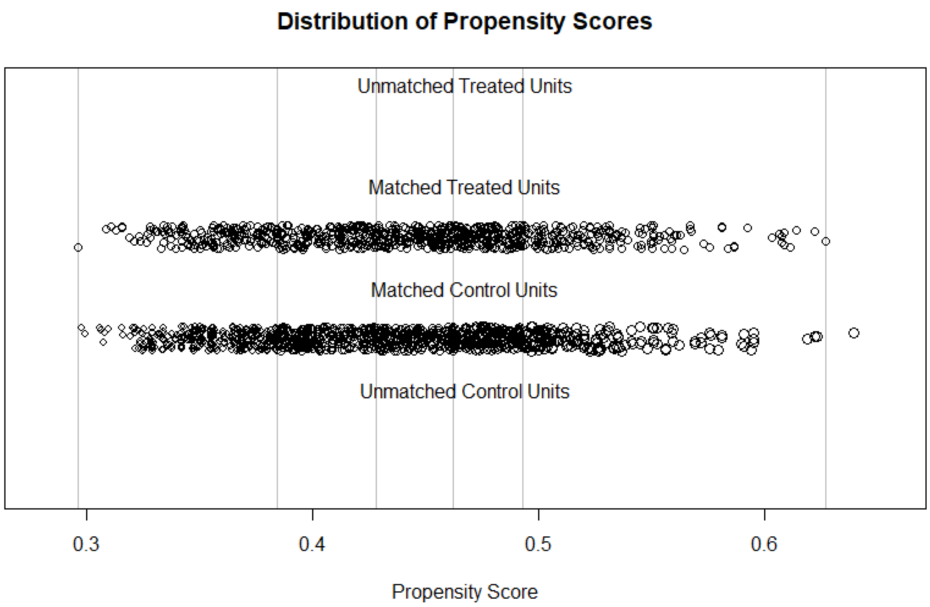


Figure D.15: Propensity Score Histograms - Limit breaking (youth)

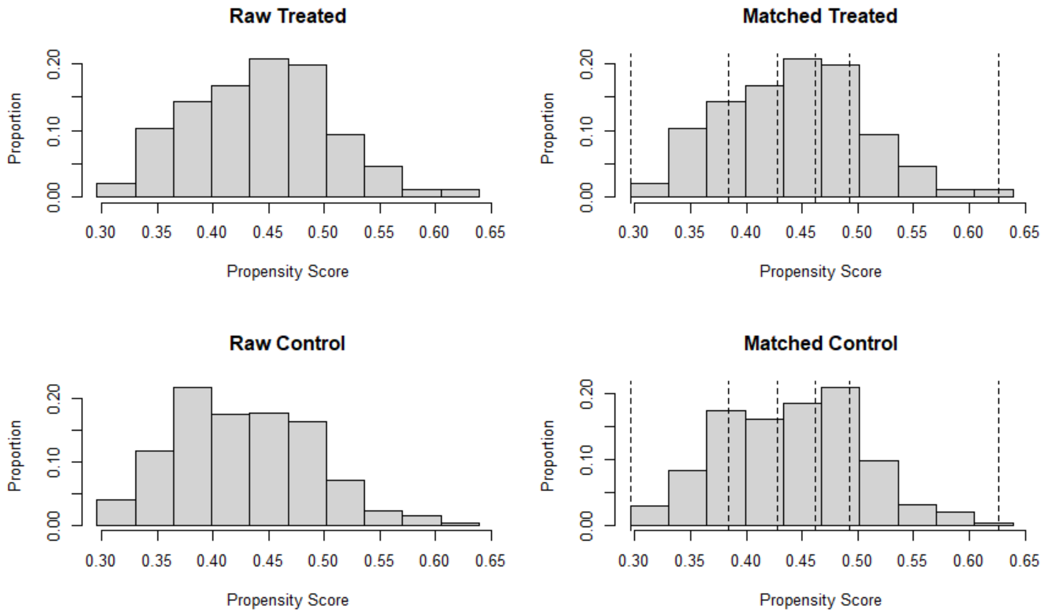


Figure D.16: Love plot - Limit breaking (parent)

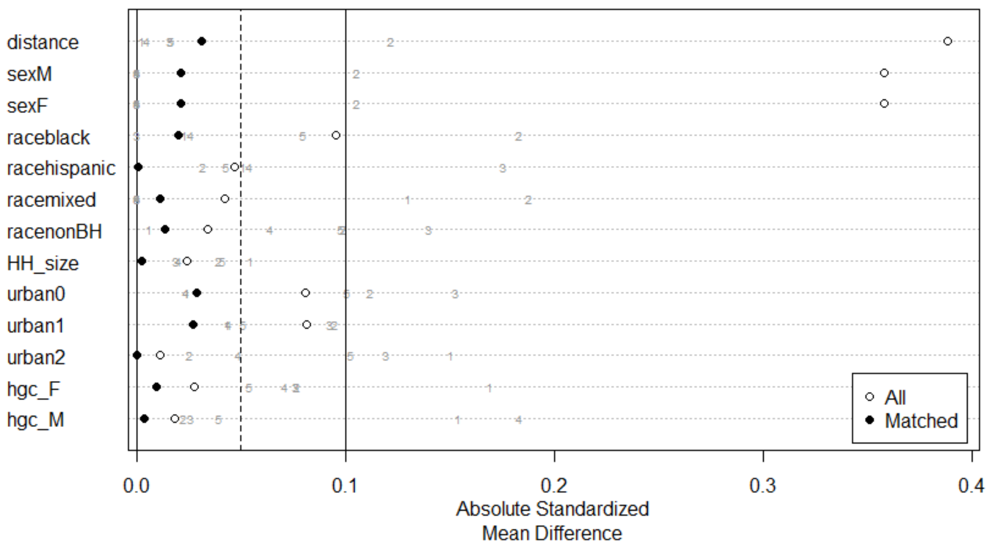


Figure D.17: Limit breaking (parent)

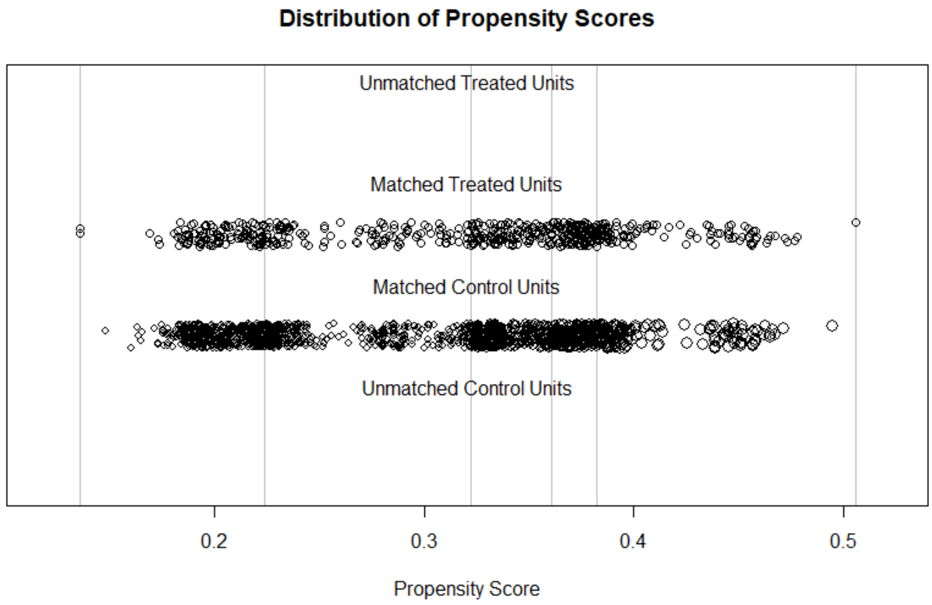
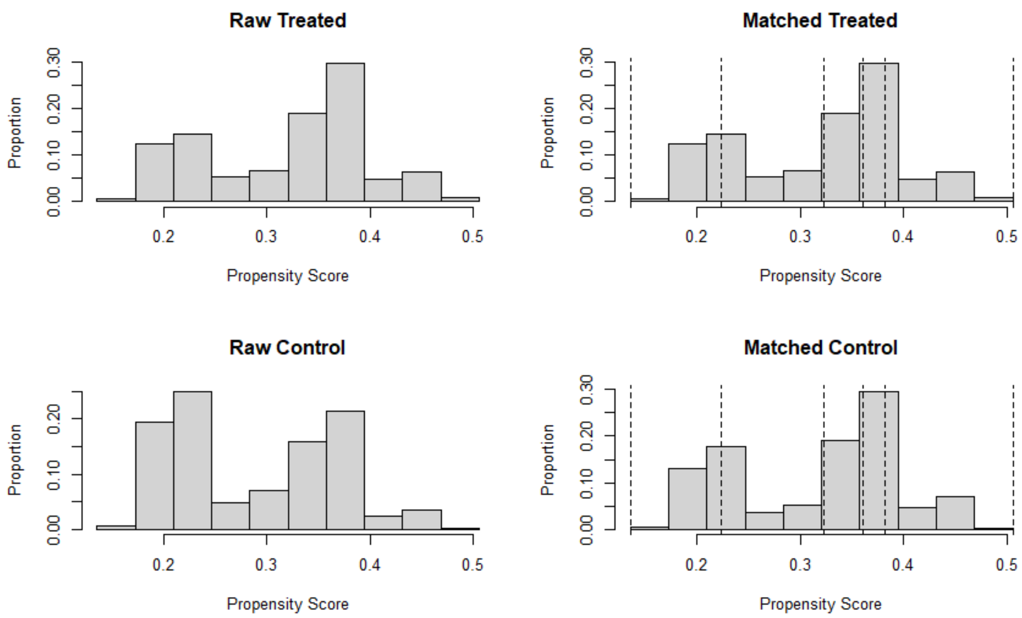


Figure D.18: Propensity Score Histograms - Limit breaking (parent)



Appendix E

Logit models

Table E.1: Subclassification

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
mon_M1	0.52^{***}					
	(0.15)					
mon_F1		0.28^{*}				
		(0.13)				
lim_Y1			-0.09			
			(0.13)			
lim_P1				-0.30		
				(0.17)		
broke_Y1					-0.49^{***}	
					(0.13)	
broke_P1						-0.52^{***}
						(0.14)
(Intercept)	1.34^{***}	1.62^{***}	1.64^{***}	1.89^{***}	1.87^{***}	1.83^{***}
	(0.12)	(0.09)	(0.09)	(0.14)	(0.09)	(0.08)
Deviance	1567.66	1551.86	1691.44	1641.47	1645.03	1633.51
Num. obs.	1868	1868	1868	1868	1868	1868

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Robust standard errors are provided in parentheses

Note: mon_M - monitoring by mother; mon_F - monitoring by father; lim_Y - limit setting reported by youth; lim_P - limit setting reported by parent; broke_Y - limit breaking reported by youth; broke_P - limit breaking reported by parent.

Table E.2: Logit model

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
mon_M	0.48^{**}					
	(0.15)					
mon_F		0.32^{*}				
		(0.14)				
lim_Y			-0.11			
			(0.13)			
lim_P				-0.38^{*}		
				(0.17)		
broke_Y1					-0.50^{***}	
					(0.13)	
broke_P1						-0.55^{***}
						(0.14)
sexF	0.51^{***}	0.55^{***}	0.54^{***}	0.53^{***}	0.50^{***}	0.45^{**}
	(0.14)	(0.14)	(0.14)	(0.14)	(0.14)	(0.14)
racehispanic	0.54^{*}	0.55^{*}	0.59^{*}	0.57^{*}	0.59^{**}	0.53^{*}
	(0.23)	(0.23)	(0.23)	(0.23)	(0.23)	(0.23)
racemixed	-0.39	-0.42	-0.46	-0.46	-0.46	-0.49
	(0.83)	(0.83)	(0.82)	(0.82)	(0.83)	(0.83)
racenonBH	0.17	0.19	0.21	0.19	0.20	0.19
	(0.18)	(0.18)	(0.18)	(0.18)	(0.18)	(0.18)
HH_size	-0.00	-0.00	0.00	0.00	0.01	0.00
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
urban1	0.01	0.04	-0.01	-0.01	0.01	0.03
	(0.16)	(0.16)	(0.16)	(0.16)	(0.16)	(0.16)
urban2	-0.14	-0.16	-0.21	-0.22	-0.20	-0.18
	(0.32)	(0.32)	(0.32)	(0.32)	(0.32)	(0.32)
hgc_F	0.14^{***}	0.14^{***}	0.14^{***}	0.14^{***}	0.14^{***}	0.14^{***}
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
hgc_M	0.12^{***}	0.13^{***}	0.12^{***}	0.13^{***}	0.12^{***}	0.13^{***}
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
(Intercept)	-2.29^{***}	-2.19^{***}	-1.99^{***}	-1.75^{***}	-1.82^{***}	-1.88^{***}
	(0.51)	(0.51)	(0.51)	(0.53)	(0.51)	(0.51)
AIC	1498.59	1503.10	1508.03	1503.48	1494.92	1493.54
BIC	1559.45	1563.96	1568.89	1564.33	1555.78	1554.40
Log Likelihood	-738.30	-740.55	-743.02	-740.74	-736.46	-735.77
Deviance	1476.59	1481.10	1486.03	1481.48	1472.92	1471.54
Num. obs.	1868	1868	1868	1868	1868	1868

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Robust standard errors are provided in parentheses

Note: mon_M - monitoring by mother; mon_F - monitoring by father; lim_Y - limit setting reported by youth; lim_P - limit setting reported by parent; broke_Y - limit breaking reported by youth; broke_P - limit breaking reported by parent; sexF - female; racehispanic - Hispanic respondents; racemixed - mixed race respondents; racenonBH - Non-Black/Non-Hispanic; HH_size - household size; urban1 - urban;, urban2 - unknown; hgc_F - highest grade completed by father; hgc_M - highest grade completed by mother.

Appendix F

Descriptive statistics

Table F.1: Mean by sex

Sex	Mon M	Mon F	Lim Y	Lim P	Broke Y	Broke P
Male	10.5	8.7	3.5	4.4	0.5	0.4
Female	11.0	8.3	3.3	4.3	0.4	0.2
Respondents who completed HS						
Male	10.7	8.9	3.5	4.3	0.4	0.3
Female	11.1	8.4	3.3	4.3	0.4	0.2
Respondents who did not complete HS						
Male	9.6	7.5	3.7	4.5	0.6	0.5
Female	10.8	7.9	3.4	4.6	0.5	0.2

Note: Mon M - monitoring by mother; Mon F - monitoring by father; Lim Y - limit setting reported by youth; Lim P - limit setting reported by parent; Broke Y - limit breaking reported by youth; Broke P - limit breaking reported by parent.

Table F.2: Means and SD by sex - males

	Monitored by mother	Not monitored by mother
HH size	4.657220	4.753191
	(1.184890)	(1.361380)
hgc M	13.19433	12.47660
	(2.943618)	(2.976131)
hgc F	13.13090	12.06809
	(3.237930)	(3.303010)
	Monitored by father	Not monitored by father
HH size	4.696035	4.666667
	(1.147475)	(1.318951)
hgc M	13.21264	12.80176
	(2.950453)	(2.971594)
hgc F	13.23563	12.46035
	(3.410500)	(3.083384)
	Limit setting (Y)	NO Limit setting (Y)
HH size	4.739496	4.624000
	(1.287754)	(1.170198)
hgc M	12.72059	13.30800
	(2.962871)	(2.943206)
hgc F	12.59454	13.14200
	(3.244021)	(3.302067)
	Limit setting (P)	NO Limit setting (P)
HH size	4.714286	4.568282
	(1.275466)	(1.059495)
hgc M	12.93458	13.30837
	(2.958723)	(2.964540)
hgc F	12.75567	13.26872
	(3.248259)	(3.375214)
	Limit breaking (Y)	NO Limit breaking (Y)
HH size	4.709677	4.653620
	(1.231594)	(1.228494)
hgc M	12.89892	13.13307
	(3.090673)	(2.846033)
hgc F	12.70968	13.02544
	(3.327591)	(3.239060)
	Limit breaking (P)	NO Limit breaking (P)
HH size	4.720670	4.656958
	(1.311517)	(1.180113)
hgc M	12.89665	13.09385
	(2.887807)	(3.010128)
hgc F	13.03236	12.60335
	(3.061314)	(3.398291)

Note: HH size - household size (range: <2;14>); HGC M - highest degree completed by mothers (range: <2;20>); HGC F - highest degree completed by father (range: <2;20>).

Table F.3: Means and SD by sex - females

	Monitored by mother	Not monitored by mother
HH size	4.728790	4.890173
	(1.341492)	(1.522913)
hgc M	12.92211	12.04624
	(2.925643)	(3.145345)
hgc F	12.94854	12.31214
	(3.212765)	(3.407537)
	Monitored by father	Not monitored by father
HH size	4.739421	4.781038
	(1.457567)	(1.299826)
hgc M	12.37940	13.05263
	(2.928987)	(3.037862)
hgc F	13.18486	12.46050
	(3.209103)	(3.272696)
	Limit setting (Y)	NO Limit setting (Y)
HH size	4.919598	4.631579
	(1.6492176)	(1.5308879)
hgc M	12.472973	13.034483
	(2.991049)	(2.954225)
hgc F	12.39698	13.17004
	(3.064905)	(3.371434)
	Limit setting (P)	NO Limit setting (P)
HH size	4.786787	4.681416
	(1.377436)	(1.384135)
hgc M	12.66216	13.01770
	(2.905785)	(3.209658)
hgc F	12.73273	13.09735
	(3.256094)	(3.260442)
	Limit breaking (Y)	NO Limit breaking (Y)
HH size	4.902017	4.669725
	(1.467045)	(1.313498)
hgc M	12.44669	12.94679
	(3.173119)	(2.849618)
hgc F	12.59366	12.97248
	(3.548877)	(3.054826)
	Limit breaking (P)	NO Limit breaking (P)
HH size	4.778894	4.754690
	(1.487781)	(1.347390)
hgc M	12.98492	12.68543
	(2.584877)	(3.092370)
hgc F	13.13065	12.73737
	(3.141931)	(3.289105)

Note: HH size - household size (range: <2;14>); HGC M - highest degree completed by mothers (range: <2;20>); HGC F - highest degree completed by father (range: <2;20>).