ABSTRACT

Title of dissertation:	ESSAYS ON SKILLS AND VICTIMIZATION
	Miguel Sarzosa, Doctor of Philosophy, 2015
Dissertation directed by:	Professor Sergio Urzúa Department of Economics

Recent literature has shown that skills are not only essential for the development of successful adults, but also that they are malleable and prone to be affected by many experiences. In this dissertation, I explore these two sides of skills and development. I use skills as a vehicle to study the consequences victimization events have on adult outcomes, and as the channels through which the gaps in those adult outcomes materialize. I use novel longitudinal surveys and rely on an empirical strategy that recognizes skills as latent constructs. First, I estimate the treatment effects being bullied and being a bully in middle school have on several outcomes measured at age 18. I find that both, victims and bullies, have negative consequences later in life. However, they differ in how non-cognitive and cognitive skills palliate or exacerbate these consequences. Then, I move on to investigate the channels that open the gaps in adult outcomes between victims and non-victims. I estimate a structural dynamic model of skill accumulation. I allow skill formation to depend on past levels of skills, parental investment and bullying. Also, I allow bullying itself to depend on each student's past skills and those of his or her classmates. I find that being bullied at age 14 depletes current skill levels by 14% of a standard deviation for the average child.

This skill depletion causes the individual to become 25% more likely to experience bullying again at age 15. Therefore bullying triggers a self-reinforcing mechanism that opens an ever-growing skill gap that reaches about one standard deviation by age 16. Finally, under the light of skills, I explore how other type of victimization, namely discrimination against sexual minorities, creates income gaps against non-heterosexual workers. I estimate a structural model that relies on the identification of unobserved skills to allow schooling choices, occupational choices and labor market outcomes to be endogenously determined and affected by the sexual preference of the worker. The results show that difference in skills, observable characteristics, and tastes for tertiary education and type of occupation, contribute to at least half of the income gaps non-heterosexuals face.

ESSAYS ON SKILLS AND VICTIMIZATION

by

Miguel Alonso Sarzosa Castillo

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2015

Advisory Committee:
Professor Sergio Urzúa, Chair
Professor Natasha Cabrera
Professor Sebastián Galiani, Advisor
Professor Soohyung Lee
Professor Andrew Sweeting

ACKNOWLEDGEMENTS

This dissertation and, in general, my acquiring of the Ph.D. degree would have been absolutely impossible without Natalia Rojas. Listing all the ways her presence in my life has been critical to my graduate school success would take up a document longer than this dissertation.

I would like to thank my advisor, Professor Sergio Urzúa for his invaluable support and brilliant guidance. His intelligence, patience and availability are extraordinary. I have always felt honored by his trust in my work, opinion and judgement. Without his trust and counsel, I would had never embarked in the challenging research projects I put forward in this dissertation.

I would also like to thank the the Economics Department faculty that believed in me during my second year of graduate school. Especially, Professor Raymond Guiteras, Professor Jessica Goldberg and Professor Sergio Urzúa. I will always be grateful for their intercession and hope to honor their vote of confidence.

I would also like to thank John Ham, John Shea, Sebastián Galiani, Soohyung Lee, Andrew Sweeting and Brian Quistorff for their comments to many parts of this dissertation. I am also indebted to Maria Fernanda Prada and Ricardo Espinoza not only for their time and friendship, but also for their comments to the code used in this research project. I would like to thank Koji Miyamoto for granting access to the data used trough out this dissertation.

I am also indebted to the people who gave me the strength to keep going through this demanding experience. I owe gratitude to Ricardo, Ignacio, Camilo, Paz, Cesar, Tatiana, Viviana, Eugenia, Romina, Ivan and Cristina, for their support, affection and laughter. Finally, I am deeply grateful with my entire family who cheered me on all the way, even in moments when positive outcomes seemed improbable. Although they are far away, I feel them by my side every day of my life.

DEDICATION

Para Natalia, Miguel, Socorro y Daniela quienes confiaron ciegamente en mi cuando parecía el fin del camino.

Para Emiliana a quien espero con impaciencia.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. BULLYING IN TEENAGERS, *THE ROLE OF COGNITIVE AND NON-COGNITIVE SKILLS*[1]

1.1 Introduction

Bullying is a behavioral phenomenon that has been placed under the spotlight in many parts of the world in recent years. Bullying is by no means new, but it has increasingly become a behavioral issue among young people. Frequent cases of suicide in school and college aged kids around the world keep reminding society of the perils and immense costs that bullying victims—and communities in general—have to bear. This phenomenon has been reinforced by the availability of new communication technologies, specially among young people, who are often the most permeable by these new technologies. Crime and bullying are among those societal features than can potentially be transformed by the flood of information and connectivity. Cybercrime and cyberbullying are therefore behavior alternatives that were not available to previous generations.

Psychologists have defined a bullying victim as a person that is repeatedly and intentionally exposed to injury or discomfort by others (Olweus, 1997). Injury or discomfort can be caused by violent contact, by insulting, by communicating private or inaccurate information and other unpleasant gestures like the exclusion from a

[1] THIS CHAPTER WAS AUTHORED JOINTLY WITH SERGIO URZÚA

group. Olweus (1997) indicates that bullying happens in environments where there are imbalances of power, and Faris and Felmlee (2011) suggest that bullying thrive in contexts where there is the need to show peer group status. Not surprisingly, schools are the perfect setting for bullying. The combination of peer pressure with the multidimensional heterogeneity of students together with a sense of self-control still not fully developed, makes schools a petri dish for bullying. Furthermore, cyberbullying has escalated this recipe for disaster to new levels. The ability to instantly and widely convey information to peers facilitates harassment. In the virtual world, the victim can be attacked from multiple directions leaving him or her defenseless. In addition, cyberbullying can sometimes provide anonymity to the perpetrators, giving an advantage that allows them to be even more hurtful.

Bullying is not only widespread, but very costly.[2] According to government statistics from stopbullying.gov, 160,000 children miss school every day in the US because of fear of being bullied (15% of those who do not show up to school every day); one of every ten students drops out or changes school because of bullying; homicide perpetrators were found to be twice as likely as homicide victims to have been bullied previously by their peers; Bully victims are between 2 to 9 times more likely to consider suicide than non-victims. In the UK, at least half of suicides among young people are related to bullying. In South Korea, one school-aged kid (10 to 19) commits suicide each day, and the suicide is the largest cause of death for people between 15 and 24 (there are 13 suicides per 100,000 people).[3]

Surprisingly, economic literature has remained aside from the research efforts that try to understand better the bullying and cybercrime phenomena. This and the next

[2] Anti-bullying campaigns and laws have been implemented in the US, Canada, UK, Germany, Scandinavia, Colombia and South Korea

[3] Suicide is South Korea single highest in the world with 31.7 suicides per 100,000 people.

Chapter of this dissertation intends to end this absence by assessing the determinants and consequences of being bullied and being a bully. Chapters 1 and 2 of this dissertation use South Korean data on teenagers. In this particular chapter, we assess the extent to which cognitive and non-cognitive skills are able to deter the occurrence of these unwanted behaviors, and also how they palliate or exacerbate the effects of bullying on several outcomes of interest like depression, life satisfaction, the incidence of smoking, drinking, some health indicators and the ability to cope with stressful situations.[4]

We use a structural model that relies in the identification of latent skills to deal with selection. Our model is flexible enough to incorporate several desirable features. First, it recognizes that cognitive and non-cognitive measures observed by the researcher are only approximations or functions of the true latent skills (Heckman et al., 2006a). Second, the model uses mixture of normals in the estimation of the distributions of the latent factors. This guarantees the flexibility required to appropriately recreate the unobserved distributions in the estimation. Therefore, we do not assume any functional form for the distributions of the factors; instead, we estimate these distributions directly and often find that they are far from normal. Third, the model does not assume linearity in the estimation. In fact, simulations show that the estimated effects of skills on the outcomes evaluated are very non-linear. Finally, the structural model allows the researcher to simulate counterfactuals for each skill level, which are used to calculate treatment effect parameters of bullying on several

[4] *Cognitive skills* are defined as "all forms of knowing and awareness such as perceiving, conceiving, remembering, reasoning, judging, imagining, and problem solving" (APA, 2006), and *non-cognitive skills* are defined as personality and motivational traits that determine the way individuals think, feel and behave (Borghans et al., 2008). Literature has shown that cognitive and non-cognitive skills are critical to the development of successful lives (see for example Murnane et al., 1995; Cawley et al., 2001; Heckman and Rubinstein, 2001; Duckworth and Seligman, 2005; Heckman et al., 2006a; Urzua, 2008).

outcomes.

This chapter contributes to the literature in several ways. First, to the best of our knowledge, this is the first attempt to assess the determinants and consequences of bullying while dealing with the problems caused by selection into becoming a victim or a perpetrator of bullying, providing insights that can potentially motivate interventions to reduce their incidence. Second, we provide evidence on how cognitive and non-cognitive affect the likelihood of being bullied, being a bully, or being a cyber-bully, and also how these endowments deter or exacerbate the consequences of these behaviors in subsequent years. The context is special, given that the population analyzed is followed during their transition from high school to an adult life. Therefore, we see the effects victimization and skills on different outcomes during very decisive times in these young lives. Third, we are able to quantify the effect of being a bully and being bullied on several outcomes controlling for the unobserved heterogeneity caused by the latent skills.

1.2 Related Literature

Economic research on bullying is very scarce. This is the case manly because of two reasons. First and foremost, the lack of longitudinal data that inquire about bullying; and second, the fact that selection into bullying is not random. Therefore, the consequences of being bullied can be confounded by the intrinsic characteristics that made the person a victim or a perpetrator in the first place. While some economic papers have been able to use long longitudinal data, they have not been able to deal with the selection issue. To the best of our knowledge there are two papers in the economic literature that address bullying in particular. Brown and Taylor (2008) use

OLS regressions and ordered probits to look at educational attainment and wages in the UK. They find that being bullied and being a bully is correlated with lower educational attainment and in consequence with lower wages later in life. The second study is that of Eriksen et al. (2012) that uses detailed Danish data, in which they use OLS and FE regressions to find correlations between bullying and grades, pregnancy, use of psychopharmachological medication, height and weight. Although these are novel efforts, none of them deal properly with the non-randomness of the bullying "treatment".[5]

Bullying is a conduct disorder. Therefore, our work on bullies relates with that of Le et al. (2005), in which they evaluate the impacts of conduct disorders during childhood. The authors bundle bullying with several other conduct disorders like stealing, fighting, raping, damaging someone's property on purpose and conning, among others. They use an Australian sample of twins to account for the self-selection that arise from genetic and environmental reasons. Using OLS regressions, they find that kids that conduct disorders were positively correlated with dropping-out from school and unemployed later in life. As in the papers of Brown and Taylor (2008) and Eriksen et al. (2012), Le et al. (2005) are unable to deal with the endogeneity that can arise from the fact that the intrinsic unobservable characteristics that influence the conduct disorder might also be influencing the realization of the outcome variables they assess.

Psychology and sociology literatures have been more prolific in terms of descriptions of bullying as a social phenomenon. For instance, findings from (Smith et al., 2004) show that bullying victims have fewer friends, are more likely to be absent from

[5] Eriksen et al. (2012) argue that the family fixed-effects help address the selection on unobservables issue. However, in Chapter 2, I present evidence that shows that birth order influences the likelihood of being victimized at age 15.

school, and do not like break times. This literature has also found that younger kids are more likely to be bullied and that this phenomenon is more frequent among boys than among girls (Boulton and Underwood, 1993; Perry et al., 1988). Interestingly, Olweus (1997) found that school and class size are not significant determinants of the likelihood of bullying occurrence. Furthermore, Ouellet-Morin et al. (2011) showed that victims' brains have unhealthy cortisol reactions that make it difficult to cope with stressful situations.

The characterizations of the victims and perpetrators highlight the importance of controlling for non-cognitive skills throughout our analysis. According to psychological research bullied children have in general less self-esteem, and have a negative view of their situation (Björkqvist et al., 1982; Olweus, 1997). Bullies have been found to lack self-control and exceed in aggressiveness towards peers and adults (Björkqvist et al., 1982). All these analyses, although descriptive, provide a critical input in the definition of the models we use in our own work.

1.3 Data

We use the Junior High School Panel (JHSP) of the Korean Youth Panel Survey (KYP). The KYP-JHSP is a longitudinal survey that started in 2003 sampling a group of second year junior high-school students (i.e., 14 year olds). The youngsters were interviewed once a year until 2008. Thus, they were followed though high-school and into the beginning of their adult life. In particular, we are able to observe higher education choices fro those that go to college and early employment choices for those not enrolling in college.

As this is a sensitive age range regarding life-path choices, the KYP-JHSP provides interesting opportunity to understand the effects of non-cognitive skills on later decisions and behavior. The KYP-JHSP pays special interest in the life-path choices made by the surveyed population, inquiring not only about their decisions, but also about the environment surrounding their choices. Youths are often asked about their motives and the reasons that drive their decision-making process. Future goals and parental involvement in such choices are frequently elicited. The KYP-JHSP is also suitable to track non-cognitive skill dynamics given that the kids are interviewed for the first time during the beginning of their teen period. This allows the researcher to observe the evolution of skills during this critical age, and to see how the quality of the teenager's environment affects the likelihood of making good choices and avoiding risky and harmful behavior.

The sample consists of 12 regions including Seoul Metropolitan City. Children were sampled according to the proportion of the second year junior high-school students present in each region. The panel consists of 3,449 youths and their parents or guardians (see descriptive statistics in Table 1.1). Subjects were consistently interviewed in six waves.[6] Each year, information was collected in two separate questionnaires: one for the teenager, and another one for the parents or guardians.

[6] As in any longitudinal survey, attrition can an issue. By wave 2, 92% of the sample remained; by wave 3, 91% did so; by wave 4, 90%; and by wave 5, 86% remained in the sample. Appendix J presents an analysis on the attrited observations. In particular, being a bully or being a victim of bullies is not a determinant for leaving the sample.

Tab. 1.1: Descriptive Statistics

Total sample size	3,449		
Number of Females	1,724	Fathers Education:	
Proportion of urban households	78.55%	High-school	42.94%
Prop. of single-headed households	6%	4yr Coll. or above	36.56%
Median monthly income per-capita	1mill won	Mothers Education:	
Prop. of Youths in College by 19	56.65%	High-school	56.31%
Incidence of smoking by 19*	19.08%	4yr Coll. or above	19.51%
Prop. of Single-child households	8.6%		

*Incidence calculated as the proportion of people who has smoked at least once in the last year

Besides inquiring about career planning and choices, the KYP-JHSP inquires about academic performance, student effort and participation in different kinds of private tutoring. The survey also asks about time allocation, leisure activities, social relations, attachment to friends and family, participation in deviant activity, and the number of times the respondent has been victimized in different settings. In addition, the survey performs a comprehensive battery of personality questions from which measures of self-esteem, self-stigmatization, self-reliance, aggressiveness, anger, self control and satisfaction with life can be constructed.

While the youngsters are often asked about the involvement of their parents in many aspects of their life, parents and guardians answer only a short questionnaire covering household composition and their education, occupation and income.

1.3.1 The Construction of the Non-Cognitive Measures

As mentioned below in the description of our empirical strategy, our estimation of the distribution parameters of the latent non-cognitive trait uses three scores that measure socio-emotional skills. The KYP-JHSP contains a comprehensive battery of measures related to socio-emotional skills. Among them, we chose the measures of locus of control, irresponsibility and self-esteem to use in our initial estimation of the distribution of non-cognitive skills.

It should be noted that most of the socio-emotional information in the KYP-JHSP is recorded in categories that group the reactions of the respondent in bins like "strongly agree" or "disagree". In consequence, and following common practice in the literature, we constructed socio-emotional skill measures by adding categorical answers of several questions regarding the same topic. This method incorporates some degree of continuity in the scores, which is essential for our estimation procedure. The questions used can be found in Appendix I, and Table 1.2 show the descriptive statistics of the constructed measures.

Tab. 1.2: Descriptive Statistics of Non-Cognitive Measures

	Locus of Control		Irresponsibility		Self-esteem	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
All	10.679	2.142	8.288	2.403	-4.051	4.455
Males	10.835	2.182	8.310	2.397	-3.848	4.445
Females	10.524	2.091	8.267	2.409	-4.252	4.455
Attending College*	11.114	1.949	8.004	2.266	-2.913	4.103
Not Attending College*	11.166	2.007	8.124	2.347	-3.142	4.537

* Sample limited to wave 6

1.3.2 The Construction of the Cognitive Measures

Although the KYP-JHSP has a rich battery of behavioral questions, unfortunately it is quite limited regarding cognitive measures. Ideally, we would like to have measures closely linked to cognitive ability that are expected to be orthogonal to non-cognitive measures, such as coding speed and digit recollection. However, the lack of such measures forces us to infer cognitive ability from grades and academic performance. In particular, we use the scores obtained in tests of i) math and science; ii) language (Korean) and social studies; and iii) the grade obtained in an overall test taken yearly. See Table 1.3 for the descriptive statistics of these measures.

Tab. 1.3: Descriptive Statistics of Cognitive Measures (Standardized)

	Math and Science		Language and Social Studies		Class grade in last semester	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
All	0.115	1.043	-0.002	1.066	-0.137	1.074
Males	0.255	1.044	-0.141	1.081	-0.192	1.067
Females	-0.024	1.024	0.008	1.050	-0.081	1.079
Attending College*	0.236	1.007	0.101	1.011	0.027	1.015
Not Attending College*	-0.043	1.068	-0.138	1.119	-0.351	1.110

* Sample limited to wave 6

Previous literature has shown that academic performance is not orthogonal to socio-emotional skills. In other words, the production function of academic test scores has to be modeled using both cognitive and socio-emotional skills as inputs. As will be shown in Section 1.5, our model takes fully into account this feature of the data and incorporates it into the estimation.

1.4 *Reduced Form Regressions and the Issue of Selection*

We would like to inquire about the effect of bulling D at time t on outcomes of interest Y measured at time $t+h$. The outcomes we consider are depression, the likelihood of smoking, drinking, life satisfaction, self reported physical and mental health, and indexes of stress on treatment variables like being bullied or being a bully.[7] Therefore the model to estimate as in Brown and Taylor (2008) and Eriksen et al. (2012) is one of the form

$$Y_{t+h} = \mathbf{X}_Y \beta + \gamma D_t + e_{t+h}$$

[7] Self reported physical health is measured as whether the respondent considers she is in good health or not. The mental health outcome is measured as whether the respondent has been diagnosed to have psychological or mental problems. Regarding the stress measures, we use two outcome measures. First, we use an index that quantifies how stressed the respondent gets regarding self image. Second, we use and index that results from the summation of the stress indexes related to stress caused by parents, friends, image and poverty. Descriptive statistics of the outcome variables can be found in Table G.1 and G.2.

where $\mathbf{X}_Y$ is a matrix with all observable controls. However, $D \not\perp e$, and therefore, $\hat{\gamma}$ will be biased. If we consider that ability or skill endowments play a role in this endogeneity, we would like to introduce measures $\mathbf{T}$ of these endowments (i.e., test scores) as controls. In order to deter reverse causality, we would like these measures to be taken before the bullying episode occurs, that is at time $t-1$. The regression equation becomes:

$$Y_{t+h} = \mathbf{X}_Y\beta + \gamma D_t + \pi\mathbf{T}_{t-1} + \nu_{t+h} \tag{1.1}$$

Tables 1.4 to 1.8 show the results of regressions of the form of (1.1). The reduced form regressions indicate that there are correlations between being bullied at 15 and depression, the likelihood of being sick, having mental health issues and feeling stressed at 18. In the same way, they show that being a bully at age 15 is correlated with being depressed, smoking and feeling stressed by 18. We see no correlations between being bullied or being a bully with drinking, life satisfaction or going to college. Tables 1.4 to 1.8 also show a interesting characteristic of these reduced form regressions: out of the the six proxies of abilities we use (i.e., locus of control, irresponsibility, self-esteem, math score, language score, class score) no more than three of them turn out to correlate with the outcome variables given the other controls in every regression.

We are not able to claim causality with these regressions because we have the strong conviction that even after controlling for test scores D is still endogenous. In addition, the components of $\mathbf{T}$ are only proxies of ability. Therefore, as we will confirm below, $\mathbf{T}$ introduces a measurement error that correlates with ν_{t+h} in (1.1). Furthermore, D and $\mathbf{T}$ are also correlated. Hence, regressions like (1.1) are problematic due to the fact that the selection and endogeneity problems are not solved by the introduction of more controls. The obvious alternatives to these issue are the use of instrumental variables and structural modeling. We choose to do the latter as we lack of any exogenous variation that we could exploit as an instrument.

Tab. 1.4: Reduced Form Regressions

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES		Depression			Drinking	
Bullied	0.430		0.330	0.006		0.003
	(0.282)		(0.286)	(0.031)		(0.031)
Bully		0.735**	0.673**		0.029	0.026
		(0.325)	(0.329)		(0.036)	(0.036)
Age in Months	-0.019	-0.028	-0.016	-0.005*	-0.004	-0.005
	(0.025)	(0.025)	(0.025)	(0.003)	(0.003)	(0.003)
Male	-0.951***	-0.937***	-0.963***	0.133***	0.131***	0.133***
	(0.187)	(0.183)	(0.187)	(0.021)	(0.020)	(0.021)
Oldersiblings	-0.239	-0.237	-0.227	0.019	0.027	0.019
	(0.177)	(0.173)	(0.177)	(0.020)	(0.019)	(0.020)
Youngsiblings	-0.133	-0.103	-0.122	-0.006	-0.003	-0.005
	(0.178)	(0.175)	(0.178)	(0.020)	(0.019)	(0.020)
Lnmonthinc_pc	-0.294	-0.251	-0.293	0.003	-0.001	0.003
	(0.187)	(0.184)	(0.187)	(0.021)	(0.020)	(0.021)
Urban	-0.254	-0.303	-0.240	-0.046	-0.050*	-0.045
	(0.271)	(0.264)	(0.270)	(0.030)	(0.029)	(0.030)
LivesBothParents	-0.826*	-0.827*	-0.828*	-0.042	-0.051	-0.042
	(0.465)	(0.451)	(0.465)	(0.051)	(0.049)	(0.051)
OnlyMother	-1.002	-0.928	-0.974	0.003	-0.003	0.004
	(0.614)	(0.600)	(0.614)	(0.068)	(0.066)	(0.068)
FatherEd2yColl	0.083	0.002	0.078	-0.023	-0.023	-0.023
	(0.349)	(0.343)	(0.349)	(0.039)	(0.038)	(0.039)
FatherEd4yColl	0.092	0.057	0.092	-0.051**	-0.051**	-0.051**
	(0.218)	(0.215)	(0.218)	(0.024)	(0.024)	(0.024)
FathereducGS	0.370	0.314	0.373	-0.011	-0.015	-0.011
	(0.398)	(0.388)	(0.398)	(0.044)	(0.042)	(0.044)
Locus of Control	-0.097	-0.087	-0.104	0.010	0.008	0.010
	(0.098)	(0.095)	(0.098)	(0.011)	(0.010)	(0.011)
Irresponsibility	0.480***	0.447***	0.459***	0.044***	0.045***	0.043***
	(0.096)	(0.095)	(0.097)	(0.011)	(0.010)	(0.011)
Self-Esteem	-0.981***	-0.994***	-0.983***	-0.035***	-0.031***	-0.035***
	(0.096)	(0.094)	(0.096)	(0.011)	(0.010)	(0.011)
Math - Science	-0.179	-0.146	-0.169	-0.003	0.001	-0.003
	(0.127)	(0.125)	(0.127)	(0.014)	(0.014)	(0.014)
Language - SSt.	0.054	0.022	0.045	0.015	0.013	0.015
	(0.131)	(0.129)	(0.131)	(0.014)	(0.014)	(0.015)
Class Score	0.270*	0.271**	0.267*	-0.044***	-0.044***	-0.044***
	(0.141)	(0.138)	(0.141)	(0.016)	(0.015)	(0.016)
Observations	2,538	2,636	2,538	2,544	2,642	2,544
R-squared	0.091	0.091	0.093	0.050	0.051	0.051

Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Intercepts of regressions not shown.

Tab. 1.5: Reduced Form Regressions

VARIABLES	(7)	(8) Smoking	(9)	(10)	(11) Life Satisfaction	(12)
Bullied	0.010		-0.006	-0.038		-0.043
	(0.022)		(0.022)	(0.031)		(0.031)
Bully		0.109***	0.113***		0.030	0.035
		(0.025)	(0.025)		(0.035)	(0.036)
Age in Months	-0.000	-0.000	0.000	0.003	0.005*	0.003
	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
Male	0.128***	0.127***	0.126***	0.035*	0.031	0.035*
	(0.014)	(0.014)	(0.014)	(0.020)	(0.020)	(0.020)
Oldersiblings	-0.014	-0.011	-0.012	0.043**	0.044**	0.044**
	(0.014)	(0.013)	(0.013)	(0.019)	(0.019)	(0.019)
Youngsiblings	-0.021	-0.017	-0.019	0.013	0.017	0.014
	(0.014)	(0.013)	(0.014)	(0.019)	(0.019)	(0.019)
Lnmonthinc_pc	-0.005	-0.007	-0.005	0.073***	0.069***	0.074***
	(0.014)	(0.014)	(0.014)	(0.020)	(0.020)	(0.020)
Urban	-0.003	-0.019	-0.001	-0.061**	-0.055*	-0.061**
	(0.021)	(0.020)	(0.021)	(0.029)	(0.029)	(0.029)
LivesBothParents	-0.071**	-0.079**	-0.072**	0.068	0.066	0.067
	(0.036)	(0.034)	(0.035)	(0.051)	(0.049)	(0.051)
OnlyMother	-0.050	-0.048	-0.046	0.069	0.068	0.071
	(0.047)	(0.046)	(0.047)	(0.067)	(0.065)	(0.067)
FatherEd2yColl	0.014	0.020	0.013	0.085**	0.088**	0.084**
	(0.027)	(0.026)	(0.027)	(0.038)	(0.037)	(0.038)
FatherEd4yColl	-0.005	-0.009	-0.005	-0.003	0.002	-0.003
	(0.017)	(0.016)	(0.017)	(0.024)	(0.023)	(0.024)
FathereducGS	0.007	0.011	0.008	0.095**	0.111***	0.095**
	(0.030)	(0.029)	(0.030)	(0.043)	(0.042)	(0.043)
Locus of Control	0.017**	0.012	0.016**	0.027**	0.024**	0.026**
	(0.007)	(0.007)	(0.007)	(0.011)	(0.010)	(0.011)
Irresponsibility	0.040***	0.039***	0.036***	-0.011	-0.012	-0.012
	(0.007)	(0.007)	(0.007)	(0.010)	(0.010)	(0.011)
Self-Esteem	-0.008	-0.009	-0.008	0.074***	0.076***	0.074***
	(0.007)	(0.007)	(0.007)	(0.010)	(0.010)	(0.010)
Math - Science	-0.009	-0.004	-0.008	0.004	0.003	0.004
	(0.010)	(0.009)	(0.010)	(0.014)	(0.014)	(0.014)
Language - SSt.	-0.008	-0.010	-0.009	0.009	0.012	0.008
	(0.010)	(0.010)	(0.010)	(0.014)	(0.014)	(0.014)
Class Score	-0.040***	-0.041***	-0.041***	0.030**	0.028*	0.030*
	(0.011)	(0.011)	(0.011)	(0.015)	(0.015)	(0.015)
Observations	2,544	2,642	2,544	2,544	2,642	2,544
R-squared	0.086	0.096	0.094	0.069	0.069	0.069

Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Intercepts of regressions not shown.

Tab. 1.6: Reduced Form Regressions

VARIABLES	(13)	(14) Sick	(15)	(16)	(17) Psychological Problems	(18)
Bullied	0.050***		0.048***	0.039***		0.036***
	(0.016)		(0.016)	(0.013)		(0.013)
Bully		0.019	0.011		0.024*	0.018
		(0.019)	(0.019)		(0.014)	(0.015)
Age in Months	0.000	-0.001	0.000	0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Male	-0.023**	-0.020*	-0.023**	0.011	0.012	0.011
	(0.011)	(0.011)	(0.011)	(0.008)	(0.008)	(0.008)
Oldersiblings	-0.018*	-0.020**	-0.017*	0.006	0.006	0.007
	(0.010)	(0.010)	(0.010)	(0.008)	(0.008)	(0.008)
Youngsiblings	-0.010	-0.012	-0.010	0.003	0.003	0.004
	(0.010)	(0.010)	(0.010)	(0.008)	(0.008)	(0.008)
Lnmonthinc_pc	-0.018*	-0.016	-0.018*	-0.004	-0.003	-0.004
	(0.011)	(0.011)	(0.011)	(0.008)	(0.008)	(0.008)
Urban	-0.011	-0.014	-0.010	-0.013	-0.016	-0.013
	(0.016)	(0.015)	(0.016)	(0.012)	(0.012)	(0.012)
LivesBothParents	0.020	0.024	0.020	0.016	0.008	0.016
	(0.027)	(0.026)	(0.027)	(0.021)	(0.020)	(0.021)
OnlyMother	0.026	0.028	0.027	0.039	0.030	0.039
	(0.035)	(0.035)	(0.035)	(0.027)	(0.027)	(0.027)
FatherEd2yColl	-0.017	-0.020	-0.017	-0.010	-0.011	-0.010
	(0.020)	(0.020)	(0.020)	(0.016)	(0.015)	(0.016)
FatherEd4yColl	0.018	0.015	0.018	0.006	0.005	0.006
	(0.012)	(0.012)	(0.012)	(0.010)	(0.010)	(0.010)
FathereducGS	0.009	0.010	0.009	0.018	0.024	0.018
	(0.023)	(0.022)	(0.023)	(0.018)	(0.017)	(0.018)
Locus of Control	-0.002	-0.001	-0.002	-0.008*	-0.007	-0.008*
	(0.006)	(0.005)	(0.006)	(0.004)	(0.004)	(0.004)
Irresponsibility	0.011*	0.009	0.010*	0.014***	0.014***	0.014***
	(0.006)	(0.005)	(0.006)	(0.004)	(0.004)	(0.004)
Self-Esteem	-0.017***	-0.015***	-0.017***	-0.010**	-0.011***	-0.010**
	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)
Math - Science	0.007	0.003	0.007	0.008	0.008	0.009
	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)
Language - SSt.	0.006	0.006	0.006	0.005	0.005	0.005
	(0.008)	(0.007)	(0.008)	(0.006)	(0.006)	(0.006)
Class Score	-0.017**	-0.016**	-0.017**	-0.002	-0.005	-0.002
	(0.008)	(0.008)	(0.008)	(0.006)	(0.006)	(0.006)
Observations	2,544	2,642	2,544	2,544	2,642	2,544
R-squared	0.019	0.014	0.019	0.020	0.016	0.021

Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Intercepts of regressions not shown.

Tab. 1.7: Reduced Form Regressions

VARIABLES	(19)	(20)	(21)	(22)	(23)	(24)
	Stress: Image			StressTotal		
Bullied	0.664***		0.555**	2.469***		2.001***
	(0.213)		(0.216)	(0.668)		(0.676)
Bully		0.851***	0.739***		3.520***	3.168***
		(0.246)	(0.249)		(0.771)	(0.778)
Age in Months	0.013	0.014	0.016	0.048	0.043	0.061
	(0.019)	(0.019)	(0.019)	(0.060)	(0.059)	(0.060)
Male	-0.958***	-0.936***	-0.971***	-1.226***	-1.066**	-1.281***
	(0.141)	(0.139)	(0.141)	(0.442)	(0.434)	(0.441)
Oldersiblings	-0.203	-0.188	-0.191	-0.145	-0.154	-0.091
	(0.134)	(0.131)	(0.134)	(0.420)	(0.410)	(0.419)
Youngsiblings	-0.273**	-0.250*	-0.262*	-0.005	0.017	0.043
	(0.135)	(0.132)	(0.134)	(0.421)	(0.415)	(0.420)
Lnmonthinc_pc	-0.370***	-0.362***	-0.368***	-0.678	-0.601	-0.670
	(0.142)	(0.140)	(0.141)	(0.444)	(0.437)	(0.443)
Urban	0.110	0.146	0.125	0.609	0.761	0.673
	(0.205)	(0.200)	(0.204)	(0.641)	(0.626)	(0.639)
LivesBothParents	-0.699**	-0.770**	-0.701**	-1.244	-1.546	-1.251
	(0.352)	(0.341)	(0.351)	(1.102)	(1.069)	(1.098)
OnlyMother	-1.348***	-1.239***	-1.317***	-3.056**	-2.897**	-2.923**
	(0.464)	(0.454)	(0.464)	(1.455)	(1.423)	(1.450)
FatherEd2yColl	0.444*	0.429*	0.438*	0.588	0.606	0.563
	(0.264)	(0.259)	(0.264)	(0.828)	(0.813)	(0.826)
FatherEd4yColl	-0.224	-0.254	-0.224	-0.020	-0.181	-0.022
	(0.165)	(0.163)	(0.165)	(0.517)	(0.509)	(0.515)
FathereducGS	-0.961***	-0.894***	-0.958***	-1.710*	-1.641*	-1.697*
	(0.301)	(0.294)	(0.300)	(0.943)	(0.920)	(0.940)
Locus of Control	-0.105	-0.079	-0.112	-0.423*	-0.416*	-0.453**
	(0.074)	(0.072)	(0.074)	(0.231)	(0.225)	(0.231)
Irresponsibility	0.220***	0.205***	0.197***	0.614***	0.575**	0.515**
	(0.073)	(0.072)	(0.073)	(0.228)	(0.226)	(0.229)
Self-Esteem	-0.656***	-0.653***	-0.659***	-1.991***	-2.020***	-2.000***
	(0.073)	(0.071)	(0.072)	(0.227)	(0.222)	(0.227)
Math - Science	0.011	0.033	0.023	0.323	0.436	0.371
	(0.096)	(0.094)	(0.096)	(0.300)	(0.296)	(0.300)
Language - SSt.	-0.117	-0.151	-0.127	0.243	0.139	0.200
	(0.099)	(0.098)	(0.099)	(0.312)	(0.307)	(0.311)
Class Score	0.285***	0.281***	0.281***	1.633***	1.603***	1.617***
	(0.107)	(0.105)	(0.107)	(0.335)	(0.328)	(0.334)
Observations	2,538	2,636	2,538	2,538	2,636	2,538
R-squared	0.094	0.091	0.097	0.075	0.076	0.081

Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Intercepts of regressions not shown.

Tab. 1.8: Reduced For Regressions: Going to College

VARIABLES	(1)	(2)	(3)
	Goes to College		
Bullied	-0.042		-0.040
	(0.030)		(0.030)
Bully		-0.021	-0.013
		(0.035)	(0.035)
Age in Months	0.002	0.002	0.002
	(0.003)	(0.003)	(0.003)
Male	-0.127***	-0.130***	-0.127***
	(0.020)	(0.019)	(0.020)
Oldersiblings	-0.024	-0.025	-0.024
	(0.018)	(0.018)	(0.018)
Youngsiblings	-0.010	-0.009	-0.010
	(0.019)	(0.018)	(0.019)
Lnmonthinc_pc	0.010	0.013	0.010
	(0.020)	(0.020)	(0.020)
Urban	-0.039	-0.040	-0.039
	(0.029)	(0.028)	(0.029)
LivesBothParents	0.202***	0.195***	0.202***
	(0.048)	(0.047)	(0.048)
OnlyMother	0.048	0.040	0.047
	(0.064)	(0.063)	(0.064)
FatherEd2yColl	0.054	0.060	0.054
	(0.038)	(0.037)	(0.038)
FatherEd4yColl	-0.006	-0.006	-0.006
	(0.023)	(0.022)	(0.023)
FathereducGS	-0.070*	-0.063	-0.070*
	(0.041)	(0.040)	(0.041)
Locus of Control	-0.003	-0.005	-0.003
	(0.011)	(0.010)	(0.011)
Irresponsibility	-0.006	-0.007	-0.006
	(0.010)	(0.010)	(0.010)
Self-Esteem	-0.018*	-0.020**	-0.018*
	(0.010)	(0.010)	(0.010)
Math - Science	0.046***	0.046***	0.046***
	(0.014)	(0.013)	(0.014)
Language - SSt.	-0.030**	-0.030**	-0.030**
	(0.014)	(0.014)	(0.014)
Class Score	0.049***	0.051***	0.049***
	(0.015)	(0.015)	(0.015)
Constant	0.568***	0.562***	0.569***
	(0.108)	(0.106)	(0.108)
Observations	2,274	2,359	2,274
R-squared	0.061	0.062	0.061

Note: Standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

1.5 Empirical Strategy

The key feature of our empirical strategy is the way we deal with the fact that underlying cognitive and non-cognitive skills are latent rather than observable, and are in turn relevant determinants of outcomes, choices and scores. The core of the empirical strategy is the assumption of a linear production function of test scores, whose inputs include both the individual observable characteristics and the latent skill endowments.[8] The insight provided by Kotlarski (1967) allows us to use observed test scores to identify the underlying distributions from which all the realizations of latent endowments are drawn, facilitating the estimation of the complete structural model. This is the case because such distributions allow us to integrate over the unobservable skill endowments in all the outcomes, choices and scores associated with the model, while still being able to retrieve the loadings associated with the skills in every equation.

The crucial step of estimating the parameters that completely describe the distributions of the underlying factors relies on a maximum likelihood estimation (MLE) in which we use the mixture of normals in order to achieve the flexibility required to mimic the true underlying distributions of the latent skill endowments. The mixture of normals not only grants us flexibility in the type of distribution we are able to replicate, but also allows us to integrate numerically using the Gauss-Hermite quadrature, which is particularly useful for calculating $E\left[f\left(X\right)\right]$ when $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ (Judd,

[8] In fact, variance decompositions of the test scores presented in Figure H.1 in the Appendix show that latent endowments explain between 5 to 10 time more the variation of the scores than the observable characteristics. However, these figures also show that even after controlling for latent endowments more that half of the variation of the scores is still unexplained. These findings go in line with our argument against the use test scores as proxies of abilities in Section 1.4. The unexplained part of the variance of test scores will correlate with ν_{t+h} in (1.1) biasing the results of the regressions. That is why we will rather identify the latent endowments from the test scores.

1998).[9]

1.5.1 The General Setup

The structural models we implement can be described as a set of measurement systems that are linked by a factor structure. In a general setup, suppose we face the following system for each individual in the sample:

$$\mathbf{Y} = \mathbf{X}_Y \beta^Y + \alpha^{Y,A}\theta^A + \alpha^{Y,B}\theta^B + \mathbf{e}^{\mathbf{Y}} \tag{1.2}$$

where $\mathbf{Y}$ is a $M \times 1$ vector of outcome variables, $\mathbf{X}_Y$ is a matrix with all observable controls for each outcome variable, $\alpha^{\mathbf{Y,A}}$ and $\alpha^{\mathbf{Y,B}}$ are vectors that contain the factor loadings for each one of the two factors (i.e., θ^A and θ^B), and $\mathbf{e}^{\mathbf{Y}}$ is a vector of error terms with distributions $f_{e^{y_m}}(\cdot)$ for every $m = 1, \ldots, M$. We assume that $\mathbf{e}^{\mathbf{Y}} \perp \left(\theta^A, \theta^B, \mathbf{X}_{\mathbf{Y}}\right)$, and also that $e^{y_i} \perp e^{y_j}$ for $i, j = 1, \ldots, M$. Furthermore, we assume the factors θ^A and θ^B follow the distributions $f_{\theta^A}(\cdot)$ and $f_{\theta^B}(\cdot)$, respectively.

As indicated by Carneiro et al. (2003), the estimations that come from the factor structure will gain interpretability and their identification will require less restrictions if a measurement system is joined to the system described in (1.2). The purpose of this adjoined system is to identify the distributional parameters of the unobserved factors. This adjoined measurement system has the following form:

$$\mathbf{T} = \mathbf{X}_T \beta^T + \alpha^{\mathbf{T,A}}\theta^A + \alpha^{\mathbf{T,B}}\theta^B + \mathbf{e}^{\mathbf{T}} \tag{1.3}$$

where $\mathbf{T}$ is a $L \times 1$ vector of measurements (e.g., test scores), $\mathbf{X}_T$ is a matrix with

[9] The structural estimations presented in this Chapter were done using the `heterofactor` command for Stata developed by Miguel Sarzosa and Sergio Urzua. See Sarzosa and Urzua (2012).

all observable controls for each measurement, and $\alpha^{\mathbf{T},\mathbf{A}}$ and $\alpha^{\mathbf{T},\mathbf{B}}$, are the loadings of the unobserved factors. Again, we assume that $\left(\alpha^{\mathbf{T},\mathbf{A}}, \alpha^{\mathbf{T},\mathbf{B}}, \mathbf{X}_T\right) \perp \mathbf{e}^{\mathbf{T}}$, that all the elements of the $L \times 1$ vector $\mathbf{e}^{\mathbf{T}}$ are mutually independent and have associated distributions $f_{e^h}(\cdot)$ for every $h = 1, \ldots, L$. Carneiro et al. (2003) show that, in order to identify the loadings and the diagonal matrix of the variances of the factors Σ_θ, we need to use two restrictions. First, we need $\theta^A \perp \theta^B$. And second, if we let k be the number of factors we are using in the model, we need L to be at least $2k + 1$. Therefore, the presence of two factors in (1.2) implies that there should be *at least* five measures in (1.3).

From the insight provided by Kotlarski (1967), we know model (1.3), and in particular the distributional parameters that describe $f_{\theta^A}(\cdot)$ and $f_{\theta^B}(\cdot)$, are non-parametrically identified (up to one normalization). Therefore, one of the loadings of each factor should be set equal to 1, and the estimation of all the rest of the loadings should be interpreted as relative to those used as nummeraire. We estimate the model using maximum likelihood estimation (MLE). The likelihood is

$$\mathcal{L} = \prod_{i=1}^{N} \int \int \left[\begin{array}{c} f_{e^1}\left(\mathbf{X}_{T_1}, T_1, \zeta^A, \zeta^B\right) \times \ldots \\ \cdots \times f_{e^L}\left(\mathbf{X}_{T_L}, T_L, \zeta^A, \zeta^B\right) \end{array} \right] dF_{\theta^A}\left(\zeta^A\right) dF_{\theta^B}\left(\zeta^B\right)$$

where we integrate over the distributions of the factors due to their unobservable nature, obtaining $\hat{\beta}^T, \alpha^{T,A}, \alpha^{T,B}, \hat{F}_{\theta^A}(\cdot)$ and $\hat{F}_{\theta^B}(\cdot)$. Note that we do not assume any functional form for the distributions of the factors $F_{\theta^A}(\cdot)$ and $F_{\theta^B}(\cdot)$. On the contrary, we estimate them.

Having identified the distributional parameters of $F_{\theta^A}(\cdot)$ and $F_{\theta^B}(\cdot)$ from (1.3), we are able to move on to estimate model (1.2). The likelihood function in this case is

$$\mathcal{L}=\prod_{i=1}^{N}\int\int\left[\begin{array}{c} f_{e^{y_1}}\left(\mathbf{X}_{Y_1},Y_1,\zeta^A,\zeta^B\right)\times\ldots \\ \cdots\times f_{e^{y_M}}\left(\mathbf{X}_{Y_M},Y_M,\zeta^A,\zeta^B\right)\end{array}\right]d\hat{F}_{\theta^A}\left(\zeta^A\right)d\hat{F}_{\theta^B}\left(\zeta^B\right)$$

This MLE will yield $\hat{\beta}^Y$, $\alpha^{Y,A}$ and $\alpha^{Y,B}$.

Note that the two steps presented above can be joined and calculated in one likelihood of the form:

$$\mathcal{L}=\prod_{i=1}^{N}\int\int\left[\begin{array}{c} f_{e^{y_1}}\left(\mathbf{X}_{Y_1},Y_1,\zeta^A,\zeta^B\right)\times\ldots \\ \cdots\times f_{e^{y_M}}\left(\mathbf{X}_{Y_M},Y_M,\zeta^A,\zeta^B\right) \\ \times f_{e^1}\left(\mathbf{X}_{T_1},T_1,\zeta^A,\zeta^B\right)\times\ldots \\ \cdots\times f_{e^L}\left(\mathbf{X}_{T_L},T_L,\zeta^A,\zeta^B\right)\end{array}\right]dF_{\theta^A}\left(\zeta^A\right)dF_{\theta^B}\left(\zeta^B\right) \tag{1.4}$$

1.5.2 Basic Model with a Choice Variable (Roy Model)

The general setup presented above in subsection (1.5.1) can be used to introduce a special case in which there is a binary treatment D (e.g., being bullied or not) and a subsequent outcome (e.g., likelihood of depression at age 19), that is, a model of potential outcomes inspired by the Roy model (Roy, 1951; Willis and Rosen, 1979). Individuals must choose between two sectors, for example, treated and not treated. The decision is based on the following choice model:

$$D=\mathbb{1}\left[\mathbf{X}_D\beta^{Y_D}+\alpha^{Y_D,A}\theta^A+\alpha^{Y_D,B}\theta^B+e^D>0\right]$$

where $\mathbb{1}[A]$ denotes an indicator function that takes a value of 1 if A is true, $\mathbf{X}_D$ represents a set of exogenous observables, and θ^A and θ^B represent the two factors drawn from the distributions $f_{\theta^A}(\cdot)$ and $f_{\theta^B}(\cdot)$. Let Y_0, Y_1 denote an outcome of interest (e.g., the likelihood of depression) for those with $D=0$ and $D=1$ respectively (e.g., non victims and bullying victims). Then, the system of equations (1.2) will represent both potential outcomes and the choice equation. That is, $\mathbf{Y}=[Y_1,Y_0,D]^{'}$ where:

$$Y_1 = \begin{cases} \mathbf{X}_Y\beta^{Y_1} + \alpha^{Y_1,A}\theta^A + \alpha^{Y_1,B}\theta^B + e^{Y_1} & \text{if } D=1 \\ 0 & \text{if } D=0 \end{cases} \tag{1.5}$$

$$Y_0 = \begin{cases} \mathbf{X}_Y\beta^{Y_0} + \alpha^{Y_0,A}\theta^A + \alpha^{Y_0,B}\theta^B + e^{Y_0} & \text{if } D=0 \\ 0 & \text{if } D=1 \end{cases} \tag{1.6}$$

$$D = \mathbb{1}\left[\mathbf{X}_D\beta^{Y_D} + \alpha^{Y_D,A}\theta^A + \alpha^{Y_D,B}\theta^B + e^D > 0\right] \tag{1.7}$$

where $\mathbf{X}_Y$ are a set of observable variables.

Now, let T_1^{NC}, T_2^{NC}, T_3^{NC}, T_1^C, T_2^C and T_3^C denote the scores of the different tests related to non-cognitive and cognitive ability of the adjoined measurement system (1.3). Then, our full system of equations is the following:

$$Y_1 = \mathbf{X}_Y \beta^{Y_1} + \alpha^{Y_1,NC}\theta^{NC} + \alpha^{Y_1,C}\theta^C + e^{Y_1} \tag{1.8}$$

$$Y_0 = \mathbf{X}_Y \beta^{Y_0} + \alpha^{Y_0,NC}\theta^{NC} + \alpha^{Y_0,C}\theta^C + e^{Y_0} \tag{1.9}$$

$$D = \mathbf{X}_D \beta^{Y_1} + \alpha^{Y_{1,NC}}\theta^{NC} + \alpha^{Y_1,C}\theta^C + e^{Y} \tag{1.10}$$

$$T_1^{NC} = \mathbf{X}_T \beta^{T_1} + \alpha^{T_1^{NC}}\theta^{NC} + e^{T_1^{NC}} \tag{1.11}$$

$$T_2^{NC} = \mathbf{X}_T \beta^{T_2} + \alpha^{T_2^{NC}}\theta^{NC} + e^{T_2^{NC}} \tag{1.12}$$

$$T_3^{NC} = \mathbf{X}_T \beta^{T_3} + \alpha^{T_3^{NC}}\theta^{NC} + e^{T_3^{NC}} \tag{1.13}$$

$$T_1^{C} = \mathbf{X}_T \beta^{T_1} + \alpha^{T_4^{NC}}\theta^{NC} + \alpha^{T_4^C}\theta^C + e^{T_1^{C}} \tag{1.14}$$

$$T_2^{C} = \mathbf{X}_T \beta^{T_2} + \alpha^{T_5^{NC}}\theta^{NC} + \alpha^{T_5^C}\theta^C + e^{T_2^{C}} \tag{1.15}$$

$$T_3^{C} = \mathbf{X}_T \beta^{T_3} + \alpha^{T_6^{NC}}\theta^{NC} + \alpha^{T_6^C}\theta^C + e^{T_3^{C}} \tag{1.16}$$

We assume the error terms (e^{Y_1}, e^{Y_0}, e^D, $e^{T_1^{NC}}$, $e^{T_2^{NC}}$, $e^{T_3^{NC}}$, $e^{T_1^{C}}$, $e^{T_2^{C}}$, $e^{T_3^{C}}$) are independent among themselves and from (θ^C, θ^{NC}, $\mathbf{X}_Y$, $\mathbf{X}_D$, $\mathbf{X}_T$) and follow distributions $f^{Y_0}(\cdot)$, $f^{Y_1}(\cdot)$, $f^{D}(\cdot)$, $f^{T_1^{NC}}(\cdot)$, $f^{T_2^{NC}}(\cdot)$, $f^{T_3^{NC}}(\cdot)$, $f^{T_1^{C}}(\cdot)$, $f^{T_2^{C}}(\cdot)$, $f^{T_3^{C}}(\cdot)$ respectively. The production functions for test scores presented in equations (1.11) to (1.16) incorporate the fact that the KYP-JHSP provides direct measures of "pure" non-cognitive skills, providing only indirect measures of cognitive ability, in the form of academic test scores, which likely reflect both cognitive and non-cognitive skills. This is why cognitive skill is excluded from equations (1.11) to (1.13), while non-cognitive skill is not excluded from (1.14) to (1.16).

We estimate the model using maximum likelihood estimation (MLE). The likelihood we estimate is

$$\mathcal{L} = \prod_{i=1}^{N} \int \int \left[\begin{array}{c} f^{Y_0}\left(\mathbf{X}_Y, Y_0, \zeta^{NC}, \zeta^{C}\right)^{1-D} f^{Y_1}\left(\mathbf{X}_Y, Y_1, \zeta^{NC}, \zeta^{C}\right)^{D} \\ \times f^{D}\left(\mathbf{X}_D, Y_D, \zeta^{NC}, \zeta^{C}\right) f^{T_1^{NC}}\left(\mathbf{X}_{T_1}, T^{NC}, \zeta^{NC}\right) \\ \times f^{T_2^{NC}}\left(\mathbf{X}_{T_2}, T^{NC}, \zeta^{NC}\right) f^{T_3^{NC}}\left(\mathbf{X}_{T_3}, T^{NC}, \zeta^{NC}\right) \\ \times f^{T_1^{C}}\left(\mathbf{X}_{T_1}, T^{NC}, \zeta^{NC}, \zeta^{C}\right) f^{T_2^{C}}\left(\mathbf{X}_{T_2}, T^{NC}, \zeta^{NC}, \zeta^{C}\right) \\ \times f^{T_3^{C}}\left(\mathbf{X}_{T}, T^{NC}, \zeta^{NC}, \zeta^{C}\right) \end{array} \right] dF_{\theta^{NC}}\left(\zeta^{NC}\right) dF_{\theta^{C}}\left(\zeta^{C}\right) \quad (1.17)$$

In order to identify this model, in particular the parameters that describe the distributions of the latent factors, we need to do two normalizations. First, by setting $\alpha^{T_3^{NC}} = 1$ we are able to identify the remaining loadings associated with the non-cognitive skill. Having identified these loadings, we need to set $\alpha^{T_3^{C}} = 1$ in order to identify the remaining loadings associated with the cognitive skill.

1.5.3 The Case Where There is no Choice Equation

It is important to note that the system described in equations (1.2) and (1.3) and estimated using equation (1.4) can easily be modified to incorporate outcomes that are not observed only conditional on a choice such as D. Then, the outcomes vector is $\mathbf{Y} = Y$. In this case, the system becomes

$$Y = \mathbf{X}_Y \beta^Y + \alpha^{Y,NC}\theta^{NC} + \alpha^{Y,C}\theta^C + e^Y \tag{1.18}$$

$$T_1^{NC} = \mathbf{X}_T \beta^{T_1} + \alpha^{T_1^{NC}}\theta^{NC} + e^{T_1^{NC}} \tag{1.19}$$

$$T_2^{NC} = \mathbf{X}_T \beta^{T_2} + \alpha^{T_2^{NC}}\theta^{NC} + e^{T_2^{NC}}$$

$$T_3^{NC} = \mathbf{X}_T \beta^{T_3} + \alpha^{T_3^{NC}}\theta^{NC} + e^{T_3^{NC}} \tag{1.20}$$

$$T_1^{C} = \mathbf{X}_T \beta^{T_1} + \alpha^{T_1^{NC}}\theta^{NC} + \alpha^{T_1^{C}}\theta^{C} + e^{T_1^{C}} \tag{1.21}$$

$$T_2^{C} = \mathbf{X}_T \beta^{T_2} + \alpha^{T_2^{NC}}\theta^{NC} + \alpha^{T_2^{C}}\theta^{C} + e^{T_2^{C}}$$

$$T_3^{C} = \mathbf{X}_T \beta^{T_3} + \alpha^{T_3^{NC}}\theta^{NC} + \alpha^{T_3^{C}}\theta^{C} + e^{T_3^{C}} \tag{1.22}$$

And the likelihood function we will estimate is

$$\mathcal{L} = \prod_{i=1}^{N} \int\int \left[\begin{array}{c} f^Y\left(\mathbf{X}_Y, Y, \zeta^{NC}, \zeta^C\right) f^{T_1^{NC}}\left(\mathbf{X}_{T_1}, T^{NC}, \zeta^{NC}\right) \\ \times f^{T_2^{NC}}\left(\mathbf{X}_{T_2}, T^{NC}, \zeta^{NC}\right) f^{T_3^{NC}}\left(\mathbf{X}_{T_3}, T^{NC}, \zeta^{NC}\right) \\ \times f^{T_1^{C}}\left(\mathbf{X}_{T_1}, T^{NC}, \zeta^{NC}, \zeta^C\right) f^{T_2^{C}}\left(\mathbf{X}_{T_2}, T^{NC}, \zeta^{NC}, \zeta^C\right) \\ \times f^{T_3^{C}}\left(\mathbf{X}_{T}, T^{NC}, \zeta^{NC}, \zeta^C\right) \end{array} \right]$$
$$dF_{\theta^{NC}}\left(\zeta^{NC}\right) dF_{\theta^{C}}\left(\zeta^{C}\right)$$

1.6 Results from the Model

1.6.1 Estimation Results

1.6.1.1 The Estimation of Non-Cognitive and Cognitive Endowments

Tables 1.9 and 1.10 show the results of our estimation of a measurement system like (1.3). In particular, we estimate equations (1.19) to (1.22) with which we identify the parameters governing the distribution function of the latent non-cognitive and cognitive factors $f_{\theta^{NC}}(\cdot)$ and $f_{\theta^{C}}(\cdot)$ as of the initial sample period ($t = 1$). We include a set of controls $\mathbf{X}_T$ representing the context surrounding the youths development. We include a gender dummy, family structure indicators, father's education attainment, monthly income per capita and the age stated in months starting from March 1989 (because all sample individuals were born within the same academic year, which goes from March to February).

Tab. 1.9: Identification of Non-cognitive Skills at $t = 1$

	Locus of Control		Irresponsibility		Self-esteem	
VARIABLES	Coeff.	Std. Err	Coeff.	Std. Err	Coeff.	Std. Err
Month of Birth	-0.014***	(0.005)	0.020***	(0.005)	-0.018***	(0.005)
male	0.134***	(0.037)	-0.025	(0.038)	0.152***	(0.036)
oldersiblings	0.024	(0.035)	-0.016	(0.035)	-0.006	(0.034)
youngsiblings	0.026	(0.035)	-0.088**	(0.036)	0.032	(0.034)
lnmonthinc_pc	0.100***	(0.034)	-0.120***	(0.035)	0.039	(0.033)
urban	0.171***	(0.056)	-0.057	(0.056)	0.020	(0.054)
livesbothprnts	0.174*	(0.106)	-0.327***	(0.108)	0.231**	(0.103)
onlymother	0.261*	(0.143)	-0.224	(0.145)	0.358**	(0.140)
fathered2yColl	0.148**	(0.074)	-0.174**	(0.075)	-0.012	(0.071)
fathered4yColl	0.209***	(0.044)	-0.200***	(0.045)	0.118***	(0.043)
fathereducGS	0.315***	(0.078)	-0.378***	(0.079)	0.167**	(0.077)
Factor Loading	0.747***	(0.112)	-0.599***	(0.086)	1	.
Constant	-0.958***	(0.181)	0.886***	(0.184)	-0.630***	(0.177)
Observations	3,109		3,109		3,109	

*** p<0.01, ** p<0.05, * p<0.1

Tab. 1.10: Identification of Cognitive Skills at $t = 1$

	Math and Science		Language & Soc. St.		Class Grade	
VARIABLES	Coeff.	Std. Err	Coeff.	Std. Err	Coeff.	Std. Err
Month of Birth	-0.011**	(0.004)	-0.011**	(0.004)	-0.017***	(0.004)
male	0.286***	(0.033)	-0.018	(0.032)	0.087***	(0.029)
oldersiblings	0.029	(0.030)	-0.045	(0.030)	-0.017	(0.027)
youngsiblings	0.078**	(0.031)	0.073**	(0.031)	0.107***	(0.028)
lnmonthinc_pc	0.164***	(0.030)	0.170***	(0.030)	0.160 ***	(0.028)
urban	0.050	(0.050)	0.097**	(0.049)	-0.026	(0.043)
livesbothprnts	0.528***	(0.094)	0.469***	(0.093)	0.472***	(0.088)
onlymother	0.578***	(0.125)	0.494***	(0.123)	0.331***	(0.116)
fathered2yColl	0.233**	(0.072)	0.151**	(0.071)	0.236***	(0.069)
fathered4yColl	0.215***	(0.038)	0.359***	(0.037)	0.299***	(0.033)
fathereducGS	0.341***	(0.070)	0.533***	(0.068)	0.442***	(0.061)
No-Cog Load	0.126***	(0.007)	0.116***	(0.007)	0.101***	(0.006)
Cogn Load	0.658***	(0.018)	0.690***	(0.018)	1	.
Constant	-1.458***	(0.158)	-1.314***	(0.156)	-1.123***	(0.144)
Observations	3,109		3,109		3,109	

*** p<0.01, ** p<0.05, * p<0.1

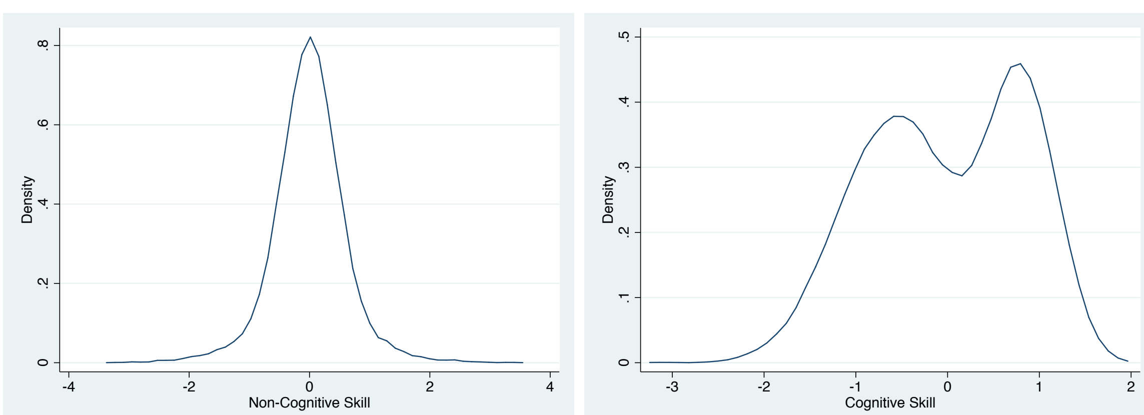
Table 1.9 present the results for the estimation of equations (1.19) to (1.20). It indicates that, as expected, non-cognitive scores do depend strongly on our estimated non-cognitive factor. That is, our estimations show that the loadings $\alpha^{T_1^{NC}}$ and $\alpha^{T_2^{NC}}$ are large and statistically different from zero at the 99% level of confidence. Our estimations of β^{T_1}, β^{T_2} and β^{T_3}, the coefficients associated with the controls contain some interesting findings. In congruence with Cunha et al. (2006) and Heckman and Masterov (2007), kids that come from wealthier, more educated parents tend to be more responsible, have higher self-control and are more positive about themselves. Our results also suggest that family composition plays a big role in fostering desirable personality traits. Kids with younger siblings and those who live with both their parents tend to be more responsible. Interestingly, the kids who live with their mother score substantially higher in the self-esteem scale than those who live only with their father. Another notable finding, which is in line with Borghans et al. (2008), is that younger kids are less responsible and have less self-control and self-esteem, even within the same year of age.

In an analogous way, Table 1.10 shows estimates of equations (1.21) to (1.22). As explained above, we allow both the cognitive and non-cognitive latent skill endowments to be inputs in the production function of our cognitive test scores. In fact, the loadings of both the cognitive and the non-cognitive latent endowments $\alpha^{T_4^{NC}}$, $\alpha^{T_4^{C}}$, $\alpha^{T_5^{NC}}$, $\alpha^{T_5^{C}}$, $\alpha^{T_6^{NC}}$ and $\alpha^{T_6^{C}}$ are highly significant in every cognitive measure used. As with the non-cognitive measures, the cognitive scores are for kids that come from wealthier and more educated parents, especially if the mother is present in the family. In addition, kids with younger siblings tend to score better in all cognitive measures, while that is not the case for those who have older siblings.

Using the estimates of equations (1.19) to (1.22) we are able to recreate the estimated distributions of initial non-cognitive and cognitive skills across the population evaluated at $t = 1$. These distributions are presented in Figures 1.1a and 1.1b.

Fig. 1.1: Distributions at $t = 1$

(a) Non-Cognitive Skills Distribution at $t = 1$ *(b)* Cognitive Skills Distribution at $t = 1$



Mean	Std. Dev	Min	Max
-.0014372	.7694455	-3.772602	4.216433

Mean	Std. Dev	Min	Max
.0006482	.9658536	-2.997454	2.1472

1.6.1.2 Bullying and Cyberbullying

We now analyze bullying in a context like the one described in subsection 1.5.3. In this setup, we evaluate the determinants of being bullied, being a bully and being a cyberbully at age 14 (i.e., $t = 2$).[10] We are particularly interested in assessing the relation between skills and these behaviors. Table 1.11 shows our results from the estimation of a model described in equations (1.18) to (1.22). The most salient finding is that while cognitive skills do no play a role in deterring or motivating any of these undesired behaviors, non-cognitive skills are a very important determinant in the likelihood of incurring in them. Our findings indicate that a one standard deviation increase in non-cognitive skills translates into a 5.69 percentage points reduction in the likelihood of being bullied. That is, an standard deviation increase in non-cognitive skills reduces by half the overall probability of being a victim of bullying. In the same vein, a standard deviation increase in non-cognitive skills reduces the likelihood of being a bully by 1.92 percentage points. This corresponds to a 24.7% reduction in the overall probability of being a bully. Finally, a standard deviation increase in non-cognitive skills tallies with a reduction of 5.08 percentage points in the likelihood of being a cyberbully, which corresponds to a 17.97% reduction in the overall probability of being a cyberbully.

The sizes of the effects are remarkable and highlight the importance of fostering non-cognitive skills to significantly reduce the incidence of bullying. Our results show that non-cognitive skills deter bullying by acting on both sides of the problem: the victims and the perpetrators. We will come back to extend this claim later in this Chapter.

[10] Recall that, as shown in the previous subsection, latent skills were measured one survey wave before (i.e., $t = 1$).

Tab. 1.11: Non-Cognitive and Cognitive Measures in $t = 1$ on Outcome Results in $t = 2$

	Physical Crimes		Cybercriminal		Being Bullied		Bully		Cyberbully	
VARIABLES	Coeff.	Std. err	Coeff.	Std. err	Coeff.	Std. err	Coeff.	Std. err	Coeff.	Std. err
Month of Birth	-0.003*	(0.002)	0.001	(0.002)	0.001	(0.002)	-0.004**	(0.001)	0.004	(0.002)
male	0.033**	(0.013)	0.045***	(0.013)	0.051***	(0.012)	0.021**	(0.011)	0.029*	(0.016)
oldersiblings	-0.027**	(0.012)	-0.013	(0.012)	-0.004	(0.011)	-0.017*	(0.010)	-0.023	(0.015)
youngsiblings	-0.032**	(0.012)	-0.020	(0.012)	-0.020*	(0.011)	-0.019*	(0.010)	-0.039**	(0.015)
lnmonthinc_pc	0.029**	(0.013)	-0.014	(0.012)	0.001	(0.012)	0.011	(0.010)	-0.013	(0.015)
urban	-0.081***	(0.019)	-0.034*	(0.019)	0.002	(0.018)	-0.041***	(0.016)	-0.020	(0.024)
livesbothparnts	-0.057*	(0.030)	0.028	(0.029)	0.003	(0.027)	-0.010	(0.024)	0.034	(0.036)
onlymother	-0.087*	(0.044)	-0.039	(0.044)	0.063	(0.041)	-0.049	(0.036)	0.055	(0.054)
fathered2yColl	0.010	(0.026)	-0.023	(0.025)	0.006	(0.024)	-0.000	(0.021)	-0.027	(0.031)
fathered4yColl	-0.007	(0.015)	-0.002	(0.015)	-0.021	(0.014)	0.002	(0.012)	-0.022	(0.019)
fathereducGS	0.016	(0.028)	0.035	(0.027)	0.017	(0.025)	-0.001	(0.023)	-0.031	(0.034)
Non-Cogn Load	-0.035**	(0.016)	-0.036**	(0.015)	-0.074***	(0.015)	-0.025*	(0.013)	-0.066***	(0.020)
Cognitive Load	-0.023***	(0.008)	0.003	(0.008)	0.002	(0.007)	-0.007	(0.006)	-0.007	(0.009)
Constant	0.180***	(0.067)	0.184***	(0.066)	0.086	(0.062)	0.121**	(0.055)	0.263***	(0.082)
Observations	2,196		2,862		2,862		2,862		2,862	

*** p<0.01, ** p<0.05, * p<0.1

In addition to the findings regarding non-cognitive endowments, the results in Table 1.11 contain other interesting findings. First, note that while age in months is associated with a higher likelihood of being a bully, it is not associated with the likelihood of being a cyberbully. This is particularly relevant because it indicates that older kids are more likely to be bullies, but not cyberbullies.[11] So, to the extent that older kids are more likely to be taller and bigger, what this might be capturing is the fact that, given the skills and other observable characteristics, physically bigger kids are more likely to be bullies and not cyberbullies. Physical size is not an issue in this case, as the attack is done through cyberspace. There is no physical confrontation and therefore physical disparities are no longer a necessary condition for bullying to happen.

Another finding related with the observable characteristics indicates that bullying is more prevalent among boys than among girls.[12] This goes in line with several woks in the psychological literature (Olweus, 1997; Wolke et al., 2001; Smith et al., 2004; Faris and Felmlee, 2011). Finally, we find that having a younger sibling deters bullying. One tentative explanation is that having a younger member of the family might give experience to the older sibling on how to deal in a positive way with people who are weaker. Therefore, while possible bullies are discouraged to take advantage of more fragile kids, possible victims might feel that they are stronger than someone, which might prevent them from accepting abuse from bullies.

[11] The coding of the Age in Months variable starts with a 1 for kids born in March 1989 and end with a 12 for the kinds born in February 1990. Therefore, older kids have lower values in the variable.

[12] Notice that due to the way gender enters the model, we are capturing its level effects on the likelihood of bullying. The analysis of gender-specific effects of skills on that likeihood is of great importance and left for future research.

1.6.1.3 The Estimation of a Model without Decision Equation

The reader should note that the adjoined measurement system described by equations (1.11) to (1.16) is the same one described in equations (1.19) to (1.22). That is, the identification process of the factor distributions is the same, and therefore, the results presented in Tables 1.9 and 1.10 hold for the models with and without decision equation.

Tables 1.12 and 1.13 show the estimations of equation (1.18). That is, without the introduction of a treatment variable. It is presented as a benchmark the see how the outcome variables are effected overall by the observable and unobservable characteristics of the respondents. These results indicate that non-cognitive skills measured at age 14 are negatively associated with the likelihood of depression, the incidence of drinking and smoking, the likelihood of being sick, having mental health issues, or feeling stressed in general at age 18. Furthermore, non-cognitive skills have a positive effect three times larger on the likelihood of having a positive perception of life than cognitive skills. This is linked with the fact that while non-cognitive skills reduce the likelihood of depression, cognitive skills increase it. Just like happens with the stress variables. However, the reduction on the likelihood of depression and stress due to non-cognitive skills is much larger than the increase in the likelihood of depression and stress caused by cognitive skills. We find no effect of cognitive skills on the incidence of drinking alcohol or having mental health issues. Finally, our results indicate that cognitive and non-cognitive skills have roughly the same effect on the reduction of the incidence of smoking.

Tab. 1.12: Non-Cognitive and Cognitive Measures in $t = 1$ on Outcome Results in $t = 6$

	Depression		Drinking		Smoking		Volunteering		Life Satisfaction	
VARIABLES	Coeff.	Std. err.	Coeff.	Std. err.	Coeff.	Std. err.	Coeff.	Std. err.	Coeff.	Std. err.
Month of Birth	0.002	(0.027)	0.001	(0.002)	0.006**	(0.002)	-0.003	(0.002)	0.006**	(0.003)
male	-1.689***	(0.192)	0.009	(0.014)	0.251***	(0.017)	-0.059***	(0.015)	0.111***	(0.021)
oldersiblings	-0.025	(0.185)	0.009	(0.014)	0.006	(0.017)	-0.016	(0.014)	0.023	(0.020)
youngsiblings	0.021	(0.183)	0.007	(0.014)	-0.003	(0.016)	-0.024*	(0.014)	0.017	(0.020)
lnmonthinc_pc	0.082	(0.206)	0.010	(0.016)	0.004	(0.019)	-0.008	(0.016)	0.027	(0.023)
urban	-0.193	(0.278)	0.044**	(0.021)	0.056**	(0.024)	0.011	(0.021)	0.026	(0.030)
liveswbothparents	-1.071**	(0.524)	-0.068*	(0.040)	-0.089*	(0.048)	0.074*	(0.041)	0.191***	(0.058)
onlymother	0.211	(0.641)	-0.074	(0.049)	-0.017	(0.058)	0.110**	(0.050)	0.089	(0.070)
fathereduc2yColl	0.146	(0.378)	0.036	(0.029)	0.012	(0.034)	0.064**	(0.029)	0.049	(0.041)
fathereduc4yColl	-0.324	(0.222)	-0.027	(0.017)	-0.039**	(0.020)	0.012	(0.017)	0.055**	(0.024)
fathereducGS	-0.145	(0.401)	-0.061**	(0.030)	-0.037	(0.036)	0.043	(0.030)	0.141***	(0.044)
Non-Cogn Loading	-2.187***	(0.239)	-0.036**	(0.019)	-0.054**	(0.021)	0.020	(0.018)	0.147***	(0.025)
Cognitive Loading	0.227**	(0.115)	0.008	(0.009)	-0.061***	(0.010)	0.014	(0.009)	0.051***	(0.012)
Constant	16.993***	(1.132)	0.843***	(0.086)	0.079	(0.102)	0.152*	(0.087)	0.098	(0.124)
Observations	2,196		2,212		2,212		2,212		2,212	

*** p<0.01, ** p<0.05, * p<0.1

Tab. 1.13: Non-Cognitive and Cognitive Measures in $t = 1$ on Outcome Results in $t = 5$

	Sick		Mental Health Prob		Stress: Image		Stress: Total	
VARIABLES	Coeff.	Std. err.	Coeff.	Std. err.	Coeff.	Std. err.	Coeff.	Std. err.
Month of Birth	-0.000	(0.001)	0.000	(0.001)	0.018	(0.019)	0.007	(0.060)
male	-0.021**	(0.010)	0.011	(0.008)	-1.045***	(0.137)	-1.306***	(0.429)
oldersiblings	-0.022**	(0.010)	0.005	(0.008)	-0.247*	(0.133)	-0.267	(0.416)
youngsiblings	-0.014	(0.010)	0.001	(0.008)	-0.347***	(0.134)	-0.095	(0.421)
lnmonthinc_pc	-0.014	(0.010)	-0.004	(0.008)	-0.414***	(0.137)	-0.445	(0.429)
urban	-0.019	(0.015)	-0.018	(0.011)	0.143	(0.195)	0.624	(0.613)
liveswbothparents	0.008	(0.022)	0.010	(0.017)	-0.802***	(0.292)	-2.134**	(0.916)
onlymother	0.018	(0.033)	0.033	(0.025)	-1.348***	(0.435)	-3.932***	(1.364)
fathereduc2yColl	-0.025	(0.020)	-0.011	(0.015)	0.446*	(0.264)	1.156	(0.831)
fathereduc4yColl	0.006	(0.012)	0.004	(0.009)	-0.417**	(0.163)	-0.168	(0.511)
fathereducGS	-0.004	(0.022)	0.017	(0.017)	-0.985***	(0.292)	-1.182	(0.915)
Non-Cogn Loading	-0.041***	(0.013)	-0.038***	(0.011)	-1.695***	(0.165)	-5.042***	(0.526)
Cognitive Loading	-0.010*	(0.006)	0.002	(0.005)	0.180**	(0.082)	1.995***	(0.257)
Constant	0.177***	(0.055)	0.048	(0.043)	13.527***	(0.732)	47.649***	(2.299)
Observations	2,667		2,667		2,667		2,660	

*** p<0.01, ** p<0.05, * p<0.1

1.6.1.4 The Estimation of a Model with Decision Equation (Roy Model)

We now turn to the estimation of the choice equation (1.10) and the outcome equations (1.9) and (1.8). Our treatments of interest are whether the youngster is bullied or not, and whether he or she is a bully or not. It should be noted that in order to facilitate identification and not rely only on functional form assumptions, we introduced an additional control in the choice equation that measures school quality.[13] As outcomes, we use the same measures used in Section 1.4.

The results are presented in Tables 1.14, 1.15, 1.16 and 1.17. Our findings indicate that skills have some differential effects on the outcomes of interest depending on whether, the person was involved in bullying or not. These findings suggests something we will return to in Subsection 1.6.2.4: skills not only influence the likelihood of being involved in bullying, but they also play a role in dealing with the negative consequences after the bullying event has occurred. For instance, cognitive skills tend to deter drinking and smoking more among victims of bullying and bullies themselves than among non-victims and non-bullies. Also, cognitive skills boost life satisfaction in a higher extent to victims than to non-victims. In the same way, non-cognitive skills reduce the likelihood of feeling stressed due to personal image more among victims than among non-victims. Non-cognitive skills increase life satisfaction among non-bullies, while that is not the case among bullies. In the same vein, non-cognitive skills are stronger preventers of mental health issues and stress among perpetrators than

[13] Th school quality measure is an index that aggregates measures of teacher responsiveness, school environment and learning conditions. The teacher responsiveness measure is based on the perceptions students have of their teacher, such as whether they think they can talk to their teacher openly and whether they would like to turn out to be like their teacher when they become adults. School environment is measured using information about robbery and criminal activity within or around the school and the presence of litter and garbage within the school or its surroundings. Finally, the learning conditions measure is based on the likelihood of students attending top institutions of higher education after graduating from that particular school, and whether students believe their school allows them to develop their talents and abilities.

among non-bullies. So regardless of whether bullying has large or small consequences on a particular dimension—which is the topic of Section 1.6.2, skill endowments help cope with these consequences in various ways depending on the outcome.

Although these findings are very informative, they do not say anything about the causal effect of bullying on later outcomes, which is the ultimate goal of implementing this model. We commit to this task in the next section.

1.6.2 Simulations

One advantage of the structural empirical strategy chosen is that we are able to recreate outcome levels as a function of the latent factors, allowing us to see simultaneously the effect of both skills on the outcome of interest. Consequently, we are able to simulate counterfactual individuals for each level of skills (Heckman et al., 2006b). This allows us to calculate the average treatment effect (ATE), the treatment effect on the treated (TT), and the treatment effect o the untreated (TUT) of being bullied or being a bully conditional on each level of cognitive and non-cognitive skills. That is $ATE\left(\theta_i^{NC}, \theta_i^{C}\right) = \mathbb{E}\left[Y_{i,1} - Y_{i,0}\left|\theta_i^{NC}, \theta_i^{C}\right.\right]$, $TT\left(\theta_i^{NC}, \theta_i^{C}\right) = \mathbb{E}\left[Y_{i,1} - Y_{i,0}\left|\theta_i^{NC}, \theta_i^{C}, D = 1\right.\right]$ and $TUT\left(\theta_i^{NC}, \theta_i^{C}\right) = \mathbb{E}\left[Y_{i,1} - Y_{i,0}\left|\theta_i^{NC}, \theta_i^{C}, D = 0\right.\right]$.

1.6.2.1 Assessing the Fit of the Model

In order to be comfortable presenting the treatment effect results drawn from simulations, we need to show that our model is able to replicate the treatments and outcomes contained in the data. Therefore, we use the results presented in Section 1.6.1 to simulate treatments and outcomes and compare them to the actual data to assess the fit of our model. First, we compare the treatment variables: being bul-

Tab. 1.14: Outcome variables at $t = 5$ After Bullied or not at $t = 2$

VARIABLES	Depression		Drinking		Smoking		Life Satisfaction	
	Coeff.	Std. err	Coeff.	Std. err	Coeff.	Std. err	Coeff.	Std. err
Bullied = 1								
Month of Birth	0.054	(0.081)	-0.004	(0.009)	-0.010	(0.007)	-0.006	(0.009)
Male	-1.915***	(0.563)	0.254***	(0.061)	0.133***	(0.046)	0.096	(0.061)
Oldersiblings	0.229	(0.523)	0.018	(0.057)	0.051	(0.042)	0.054	(0.057)
Youngsiblings	-0.165	(0.521)	0.079	(0.057)	0.072*	(0.042)	-0.014	(0.057)
lnmonthinc_pc	0.072	(0.572)	0.044	(0.062)	0.018	(0.047)	0.031	(0.063)
Urban	0.101	(0.839)	-0.061	(0.091)	0.056	(0.068)	-0.036	(0.092)
Liveswbothparents	0.235	(1.172)	0.014	(0.128)	-0.132	(0.096)	0.005	(0.128)
Onlymother	-1.517	(1.605)	0.148	(0.174)	-0.078	(0.131)	0.027	(0.175)
Fathereduc2yColl	0.394	(1.001)	-0.028	(0.109)	0.103	(0.082)	-0.065	(0.110)
Fathereduc4yColl	-0.058	(0.684)	-0.051	(0.074)	0.070	(0.056)	0.063	(0.074)
FathereducGS	0.683	(1.074)	-0.010	(0.116)	0.062	(0.087)	0.090	(0.117)
Non-Cognitive	-2.495***	(0.542)	-0.092	(0.056)	-0.017	(0.044)	0.221***	(0.056)
Cognitive	-0.300	(0.330)	-0.107***	(0.036)	-0.157***	(0.027)	0.110***	(0.036)
Constant	15.833***	(2.909)	0.232	(0.315)	0.077	(0.237)	0.384	(0.317)
Bullied = 0								
Month of Birth	-0.005	(0.028)	-0.003	(0.003)	0.003	(0.002)	0.001	(0.003)
Male	-1.117***	(0.198)	0.116***	(0.021)	0.131***	(0.015)	0.047**	(0.021)
Oldersiblings	-0.304	(0.194)	0.020	(0.021)	-0.017	(0.014)	0.043**	(0.021)
Youngsiblings	-0.181	(0.194)	-0.024	(0.021)	-0.036**	(0.014)	0.030	(0.021)
lnmonthinc_pc	-0.412**	(0.201)	-0.009	(0.022)	-0.017	(0.015)	0.095***	(0.022)
Urban	-0.329	(0.295)	-0.040	(0.032)	-0.000	(0.022)	-0.062*	(0.032)
Liveswbothparents	-1.096**	(0.536)	-0.058	(0.059)	-0.066	(0.040)	0.121**	(0.058)
Onlymother	-1.161*	(0.698)	-0.005	(0.076)	-0.049	(0.052)	0.134*	(0.075)
Fathereduc2yColl	-0.020	(0.385)	-0.031	(0.042)	-0.014	(0.028)	0.111***	(0.041)
Fathereduc4yColl	-0.125	(0.235)	-0.069***	(0.025)	-0.035**	(0.017)	0.024	(0.025)
FathereducGS	-0.021	(0.435)	-0.039	(0.047)	-0.039	(0.032)	0.143***	(0.046)
Non-Cognitive	-2.675***	(0.258)	-0.086***	(0.027)	-0.048**	(0.020)	0.186***	(0.026)
Cognitive	0.208*	(0.120)	-0.041***	(0.013)	-0.052***	(0.009)	0.039***	(0.013)
Constant	19.714***	(1.075)	0.660***	(0.117)	0.235***	(0.080)	-0.029	(0.116)
Observations	2,506		2,512		2,512		2,512	

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Tab. 1.15: Outcome variables at $t = 5$ After Bullied or not at $t = 2$

	College		Sick		psycho		StressImage		StressTotal	
VARIABLES	Coeff.	Std. err	Coeff.	Std. err	Coeff.	Std. err	Coeff.	Std. err	Coeff.	Std. err
Bullied=1										
Month of Birth	-0.000	(0.008)	0.001	(0.006)	0.008	(0.005)	0.084	(0.065)	0.035	(0.204)
Male	-0.099	(0.061)	-0.038	(0.041)	0.029	(0.035)	-0.960**	(0.454)	-0.299	(1.425)
Oldersiblings	-0.063	(0.053)	-0.037	(0.039)	0.037	(0.033)	0.438	(0.422)	1.913	(1.325)
Youngsiblings	-0.031	(0.058)	0.034	(0.038)	-0.009	(0.033)	-0.134	(0.420)	0.765	(1.318)
lnmonthinc_pc	-0.018	(0.061)	0.014	(0.042)	-0.028	(0.036)	-0.802*	(0.462)	-0.776	(1.448)
Urban	0.096	(0.083)	-0.048	(0.062)	-0.038	(0.053)	0.290	(0.677)	2.704	(2.123)
Liveswbothparents	0.537***	(0.142)	0.079	(0.087)	-0.053	(0.074)	-0.904	(0.944)	0.376	(2.977)
Onlymother	0.266	(0.177)	0.031	(0.119)	-0.159	(0.101)	-2.539**	(1.292)	-4.124	(4.065)
Fathereduc2yColl	0.024	(0.112)	-0.038	(0.074)	-0.070	(0.063)	0.951	(0.805)	0.172	(2.532)
Fathereduc4yColl	0.087	(0.072)	0.020	(0.050)	0.037	(0.043)	-0.504	(0.552)	-1.490	(1.730)
FathereducGS	-0.009	(0.113)	-0.078	(0.079)	0.083	(0.067)	-0.996	(0.865)	-3.703	(2.713)
Non-Cognitive	0.013	(0.057)	-0.051	(0.040)	-0.027	(0.034)	-2.255***	(0.410)	-6.259***	(1.332)
Cognitive	0.043	(0.036)	-0.011	(0.024)	-0.003	(0.021)	0.408	(0.267)	1.917**	(0.837)
Constant	0.238	(0.334)	0.040	(0.215)	0.180	(0.183)	14.444***	(2.348)	44.593***	(7.365)
Bullied=0										
Month of Birth	0.001	(0.003)	0.000	(0.001)	0.000	(0.001)	0.019	(0.021)	0.052	(0.064)
Male	-0.120***	(0.020)	-0.024**	(0.011)	0.007	(0.008)	-1.115***	(0.147)	-1.706***	(0.460)
Oldersiblings	-0.007	(0.020)	-0.016	(0.010)	0.004	(0.008)	-0.331**	(0.145)	-0.518	(0.452)
Youngsiblings	-0.005	(0.019)	-0.017	(0.010)	0.003	(0.008)	-0.379***	(0.145)	-0.194	(0.452)
lnmonthinc_pc	0.004	(0.020)	-0.024**	(0.011)	-0.003	(0.008)	-0.388***	(0.150)	-0.638	(0.468)
Urban	-0.028	(0.029)	-0.008	(0.016)	-0.013	(0.012)	0.043	(0.219)	0.337	(0.685)
Liveswbothparents	0.234***	(0.040)	0.010	(0.029)	0.023	(0.022)	-0.912**	(0.401)	-1.777	(1.251)
Onlymother	0.088	(0.065)	0.030	(0.038)	0.066**	(0.029)	-1.476***	(0.522)	-3.648**	(1.628)
Fathereduc2yColl	0.082**	(0.040)	-0.017	(0.021)	-0.001	(0.016)	0.437	(0.286)	1.286	(0.894)
Fathereduc4yColl	0.007	(0.024)	0.011	(0.013)	0.002	(0.010)	-0.369**	(0.175)	0.191	(0.547)
FathereducGS	-0.050	(0.043)	0.013	(0.023)	0.006	(0.018)	-1.098***	(0.323)	-1.075	(1.008)
Non-Cognitive	-0.048*	(0.026)	-0.042***	(0.014)	-0.044***	(0.012)	-1.588***	(0.190)	-4.557***	(0.601)
Cognitive	0.075***	(0.012)	-0.009	(0.006)	0.004	(0.005)	0.169*	(0.089)	2.069***	(0.277)
Constant	0.537***	(0.113)	0.194***	(0.058)	0.026	(0.044)	13.645***	(0.802)	48.097***	(2.507)
Observations			2,512		2,512		2,506		2,506	

*** p<0.01, ** p<0.05, * p<0.1

Tab. 1.16: Outcome variables at $t = 5$ After Being a Bully or not at $t = 2$

	Depression		Drinking		Smoking		Life Satisfaction	
VARIABLES	Coeff.	Std. err	Coeff.	Std. err	Coeff.	Std. err	Coeff.	Std. err
Bully=1								
Month of Birth	0.083	(0.094)	-0.014	(0.010)	-0.002	(0.009)	-0.023**	(0.009)
Male	-1.671**	(0.676)	0.050	(0.068)	0.096	(0.063)	0.214***	(0.068)
Oldersiblings	-0.015	(0.668)	0.003	(0.068)	0.012	(0.062)	-0.091	(0.068)
Youngsiblings	0.110	(0.654)	-0.173***	(0.066)	-0.014	(0.061)	-0.018	(0.066)
lnmonthinc_pc	0.451	(0.696)	-0.069	(0.071)	-0.086	(0.065)	0.067	(0.070)
Urban	-1.274	(0.940)	-0.112	(0.096)	0.018	(0.088)	0.006	(0.095)
Liveswbothparents	-1.746	(1.583)	0.156	(0.161)	-0.304**	(0.148)	0.079	(0.161)
Onlymother	-0.195	(2.507)	0.038	(0.254)	-0.364	(0.233)	-0.044	(0.253)
Fathereduc2yColl	0.182	(1.259)	0.072	(0.128)	-0.019	(0.117)	0.203	(0.127)
Fathereduc4yColl	-0.037	(0.798)	-0.039	(0.081)	-0.030	(0.074)	0.047	(0.080)
FathereducGS	2.670*	(1.487)	0.025	(0.149)	0.146	(0.137)	0.166	(0.148)
Non-Cognitive	-2.956***	(0.853)	-0.116	(0.080)	-0.142*	(0.073)	0.096	(0.080)
Cognitive	-0.507	(0.383)	-0.109***	(0.039)	-0.082**	(0.035)	0.039	(0.038)
Constant	17.145***	(3.608)	1.036***	(0.366)	0.894***	(0.336)	0.288	(0.365)
Bully=0								
Month of Birth	-0.015	(0.027)	-0.002	(0.003)	0.002	(0.002)	0.004	(0.003)
Male	-1.166***	(0.191)	0.138***	(0.021)	0.135***	(0.014)	0.034	(0.020)
Oldersiblings	-0.263	(0.185)	0.029	(0.020)	-0.009	(0.013)	0.055***	(0.020)
Youngsiblings	-0.186	(0.187)	0.005	(0.020)	-0.022	(0.014)	0.032	(0.020)
lnmonthinc_pc	-0.379**	(0.193)	-0.004	(0.021)	-0.013	(0.014)	0.083***	(0.021)
Urban	-0.260	(0.284)	-0.046	(0.031)	-0.018	(0.021)	-0.055*	(0.031)
Liveswbothparents	-0.808	(0.493)	-0.082	(0.054)	-0.071*	(0.036)	0.104*	(0.053)
Onlymother	-1.163*	(0.646)	-0.012	(0.071)	-0.037	(0.048)	0.126*	(0.070)
Fathereduc2yColl	-0.063	(0.368)	-0.038	(0.040)	0.010	(0.027)	0.089**	(0.040)
Fathereduc4yColl	-0.187	(0.228)	-0.069***	(0.025)	-0.032*	(0.017)	0.035	(0.024)
FathereducGS	-0.153	(0.409)	-0.044	(0.044)	-0.031	(0.030)	0.147***	(0.044)
Non-Cognitive	-2.688***	(0.235)	-0.079***	(0.025)	-0.042**	(0.018)	0.205***	(0.024)
Cognitive	0.229**	(0.115)	-0.043***	(0.012)	-0.060***	(0.008)	0.048***	(0.012)
Constant	19.322***	(1.023)	0.614***	(0.112)	0.216***	(0.075)	-0.004	(0.111)
Observations	2,590		2,596		2,596		2,596	

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Tab. 1.17: Outcome variables at $t = 5$ After Being a Bully or not at $t = 2$

	College		Sick		psycho		StressImage		StressTotal	
VARIABLES	Coeff.	Std. err	Coeff.	Std. err	Coeff.	Std. err	Coeff.	Std. err	Coeff.	Std. err
Bully=1										
Month of Birth	0.000	(0.009)	-0.000	(0.006)	0.004	(0.004)	0.042	(0.074)	0.054	(0.216)
Male	-0.265***	(0.067)	0.009	(0.042)	0.054*	(0.030)	-1.307**	(0.533)	-2.727*	(1.555)
Oldersiblings	-0.098	(0.062)	0.053	(0.042)	0.019	(0.027)	-0.350	(0.528)	-0.358	(1.544)
Youngsiblings	-0.099	(0.064)	0.085**	(0.041)	0.017	(0.027)	-0.380	(0.517)	-0.093	(1.511)
lnmonthinc_pc	0.017	(0.072)	0.005	(0.044)	-0.022	(0.027)	-0.295	(0.553)	0.677	(1.612)
Urban	0.130	(0.089)	0.003	(0.059)	-0.017	(0.034)	0.359	(0.742)	1.594	(2.165)
Liveswbothparents	0.251*	(0.141)	0.098	(0.100)	-0.095*	(0.053)	-0.350	(1.252)	2.353	(3.651)
Onlymother	-0.046	(0.242)	-0.008	(0.158)	-0.042	(0.078)	0.361	(1.970)	2.251	(5.753)
Fathereduc2yColl	0.021	(0.142)	0.049	(0.079)	0.011	(0.063)	0.213	(0.995)	-2.369	(2.905)
Fathereduc4yColl	-0.031	(0.074)	0.040	(0.050)	0.074**	(0.034)	0.110	(0.634)	-0.034	(1.843)
FathereducGS	-0.089	(0.141)	0.024	(0.092)	0.096*	(0.056)	-0.702	(1.164)	-0.470	(3.416)
Non-Cognitive	-0.111	(0.078)	-0.046	(0.052)	-0.326***	(0.015)	-2.642***	(0.634)	-8.592***	(1.913)
Cognitive	0.035	(0.038)	-0.007	(0.024)	0.011	(0.014)	0.041	(0.303)	0.595	(0.880)
Constant	0.511	(0.388)	-0.121	(0.227)	0.101	(0.138)	12.976***	(2.863)	41.346***	(8.353)
Bully=0										
Month of Birth	0.001	(0.003)	-0.000	(0.001)	0.001	(0.001)	0.021	(0.020)	0.036	(0.063)
Male	-0.103***	(0.020)	-0.025**	(0.011)	0.004	(0.008)	-1.068***	(0.142)	-1.362***	(0.448)
Oldersiblings	-0.007	(0.019)	-0.027***	(0.010)	0.007	(0.008)	-0.214	(0.138)	-0.215	(0.435)
Youngsiblings	0.003	(0.019)	-0.022**	(0.010)	0.002	(0.008)	-0.311**	(0.140)	-0.048	(0.439)
lnmonthinc_pc	0.004	(0.020)	-0.021*	(0.011)	-0.003	(0.008)	-0.415***	(0.145)	-0.656	(0.456)
Urban	-0.042	(0.028)	-0.016	(0.016)	-0.015	(0.012)	0.078	(0.212)	0.618	(0.668)
Liveswbothparents	0.246***	(0.039)	0.017	(0.028)	0.006	(0.021)	-1.027***	(0.371)	-2.130*	(1.165)
Onlymother	0.086	(0.061)	0.029	(0.037)	0.017	(0.028)	-1.618***	(0.485)	-4.019***	(1.525)
Fathereduc2yColl	0.084**	(0.038)	-0.029	(0.021)	-0.010	(0.016)	0.466*	(0.275)	1.290	(0.866)
Fathereduc4yColl	0.017	(0.023)	0.005	(0.013)	-0.007	(0.010)	-0.450***	(0.170)	-0.158	(0.535)
FathereducGS	-0.032	(0.041)	0.001	(0.023)	0.015	(0.017)	-1.065***	(0.305)	-1.441	(0.957)
Non-Cognitive	-0.030	(0.024)	-0.041***	(0.013)	-0.031***	(0.011)	-1.574***	(0.172)	-4.654***	(0.551)
Cognitive	0.077***	(0.012)	-0.010	(0.006)	0.003	(0.005)	0.200**	(0.086)	2.154***	(0.268)
Constant	0.523***	(0.109)	0.206***	(0.058)	0.046	(0.044)	13.697***	(0.768)	48.032***	(2.417)
Observations			2,590		2,596		2,590		2,590	

*** p<0.01, ** p<0.05, * p<0.1

lied and being a bully. Our model predicts almost perfectly the likelihood of being treated. While the data shows that 11.07% of the sample declares being bullied, our model predicts that 11.08% of the sample receives the "bullied treatment". In the same way, while the data shows that 7.77% of the sample reports being a bully, our model predicts that 7.77% of the sample receives the "bully treatment".

The next step is to assess the fit of the model in terms of the outcome variables. That is, we will compare $E[Y_0|D=0]$ and $E[Y_1|D=1]$ between the data and our model. Tables 1.18 and 1.19 show data-model comparisons of the outcome variables for the bullied treatment and the bully treatment respectively. We see that our model is able to recreate the data in a very precise way. This gives us confidence about our capability of simulation counterfactuals: $E[Y_1|D=0]$ and $E[Y_0|D=1]$, also presented in Tables 1.18 and 1.19.

Tab. 1.18: Assessing the Fit of the Model: Outcomes for Bullied

	Depression		Smoking		Sick	
	Data	Model	Data	Model	Data	Model
$E[Y_0\|D=0]$	15.5955	15.5595	.1252	.1396	.0636	.0605
$E[Y_1\|D=1]$	16.1523	16.1088	.1706	.1762	.1122	.1178
$E[Y_0\|D=1]$	■	15.9442	■	.1748	■	.0659
$E[Y_1\|D=0]$	■	15.9602	■	.1624	■	.1252

	Mental Health		StessImage		StressTotal	
	Data	Model	Data	Model	Data	Model
$E[Y_0\|D=0]$	.0369	.0324	10.0757	10.0775	42.7570	42.96
$E[Y_1\|D=1]$	.0825	.0854	10.7119	10.7525	46.4799	45.7157
$E[Y_0\|D=1]$	■	.0428	■	10.2525	■	43.6546
$E[Y_1\|D=0]$	■	.0756	■	10.4152	■	44.5614

Tab. 1.19: Assessing the Fit of the Model: Outcomes for Bully

	Depression		Smoking		Sick	
	Data	Model	Data	Model	Data	Model
$E[Y_0 \mid D=0]$	15.5926	15.5601	.1185	.1311	.0685	.0731
$E[Y_1 \mid D=1]$	16.3727	16.4359	.2364	.2637	.0904	.0989
$E[Y_0 \mid D=1]$	■	15.5032	■	.141	■	.0715
$E[Y_1 \mid D=0]$	■	16.5786	■	.2561	■	.0966
	Mental Health		StessImage		StressTotal	
	Data	Model	Data	Model	Data	Model
$E[Y_0 \mid D=0]$	.0389	.0398	10.0767	10.0819	42.9791	42.9704
$E[Y_1 \mid D=1]$	.0724	.0734	10.8909	10.8445	46.2454	46.5123
$E[Y_0 \mid D=1]$	■	.0401	■	10.0036	■	42.8792
$E[Y_1 \mid D=0]$	■	.0098	■	10.967	■	46.6772

1.6.2.2 ATE, TT and TUT of Being Bullied

Once we have estimated the counterfactuals, we are able to calculate treatment-effect parameters. Our first step is to calculate those tratment-effect parameters for the case where the treatment is being a victim of bullying. Table 1.20 shows there are significant effects of victimization on depression, sickness, mental health issues and stress. Our results indicate that being bullied at age 15 increases the chances of suffering from depression at age 18 by a tenth of a standard deviation. Being bullied also increases the chances of reporting to have health issues. In fact, bullying victimization causes the incidence of sickness almost to double. In the same way, the incidence of mental health issues is also doubled due to victimization. Regarding the stress measures, we find that being bullied increases the incidence of stress due to self-image by a standard deviation.

Tab. 1.20: ATE, TT and TUT of Being Bullied

	Depre	Smoking	Sickness	Mental	StressImage	StressTotal
ATE	0.3962***	0.0239	0.0669**	0.0421*	0.3390***	1.6620***
	(0.032479)	(0.0272)	(0.0271)	(0.0251)	(0.0308)	(0.0507)
TT	0.2274**	0.0151	0.0536	0.0437	0.3443***	1.7744***
	(0.1024)	(0.0827)	(0.0825)	(0.0826)	(0.0934)	(0.1547)
TUT	0.4197***	0.0252	.0685**	0.0420	0.3417***	1.6603***
	(0.0349)	(0.0295)	(0.0294)	(0.0272)	(0.0334)	(0.0548)

1.6.2.3 ATE, TT and TUT of Being a Bully

We now move to the calculation of treatment-effect measures for those who chose to be bullies. Table 1.21 shows that being a bully at age 15 has costs later in life. Our findings indicate that these cost are seen in term of an increase likelihood of depression, and increase in tobacco use and higher levels of stress by the age of 18. Being a bully increases the likelihood of depression by a fourth of a standard deviation, and increases the likelihood of tobacco use by more than half. As in the case of victimization, bullies are more likely to be stressed both due to their image and in general.

The reader should note that the bullying measures come from self reporting. This is key to understand why some of the negative effects are larger for bullies than for the victims. The treatments are not the same and the self-report of self-selection into them has different connotations. What a given victim considers to be bullying, might not be considered as bullying by the perpetrator. The perpetrator might consider a joke or a tease, what the victim might regard as bullying. In that sense, the "being a bully" treatment has a higher intensity than the "being bullied" treatment, and therefore the scars left for the future differ in depth.

Tab. 1.21: TE, TT and TUT of Being a Bully

	Depre	Smoking	Sickness	Mental	StressImage	StressTotal
ATE	1.0729***	0.1216***	0.0268	-0.0292	0.8905***	3.7838***
	(0.0354)	(0.0271)	(0.0271)	(0.0273)	(0.0314)	(0.0691)
TT	1.0045***	0.1175	0.0255	-0.0255	0.8693***	3.6940***
	(0.1255)	(0.0975)	(0.0973)	(0.0981)	(0.1123)	(0.2449)
TUT	1.0787***	0.1219***	0.0269	-0.0296	0.8923***	3.7914***
	(.0369)	(0.0283)	(0.0282)	(0.0285)	(0.0327)	(0.0720)

1.6.2.4 *Cognitive and Non-Cognitive Skills on the ATE of Being a Bully and Being a Bully*

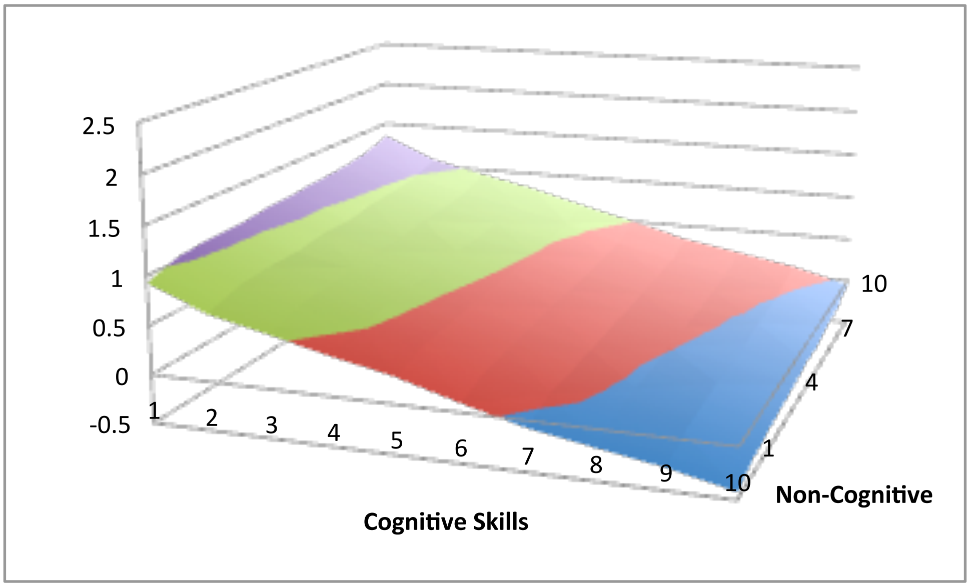
Now that we stablished the ATE of bullying on several outcomes, we can start breaking up that effect into the effects attributed to each level of latent skills. That is, we can quantify the treatment effect for individuals with particular levels of cognitive and non-cognitive skills. These results are best presented using 3D graphs with cognitive and non-cognitive skills in the x and y axes and the treatment effect in the vertical axis as in Heckman et al. (2006b) and Heckman et al. (2011b), among others.

Figures 1.2a and 1.2b show how the treatment effect of being bully and being a bully depends on the levels of skills. We see that in both cases the increase in the likelihood is a phenomenon that happens in the entire cognitive-noncognitive space. Interestingly, we see that people with lower levels of cognitive skills have a higher treatment effect for both treatments, while it is stable in the non-cognitive skill dimension.

If we now focus on the incidence if smoking, we see very interesting patterns. Figure 1.3a shows that while the effect of bullying on the victims is on average not different from zero, bullied people with lower cognitive skills tend to smoke more. On the contrary, Figure 1.3b shows that for former bullies, there is a positive effect of bullying on smoking for the entire cognitive-noncognitive space. However, this effect diminishes with higher levels of non-cognitive skills. The treatment effect of being a bully on smoking is the highest for people with low cognitive and non-cognitive skills. For this subpopulation being a bully more than doubles the likelihood of smoking.

Fig. 1.2: ATE on Depression

(a) Bullied



(b) Being a Bully

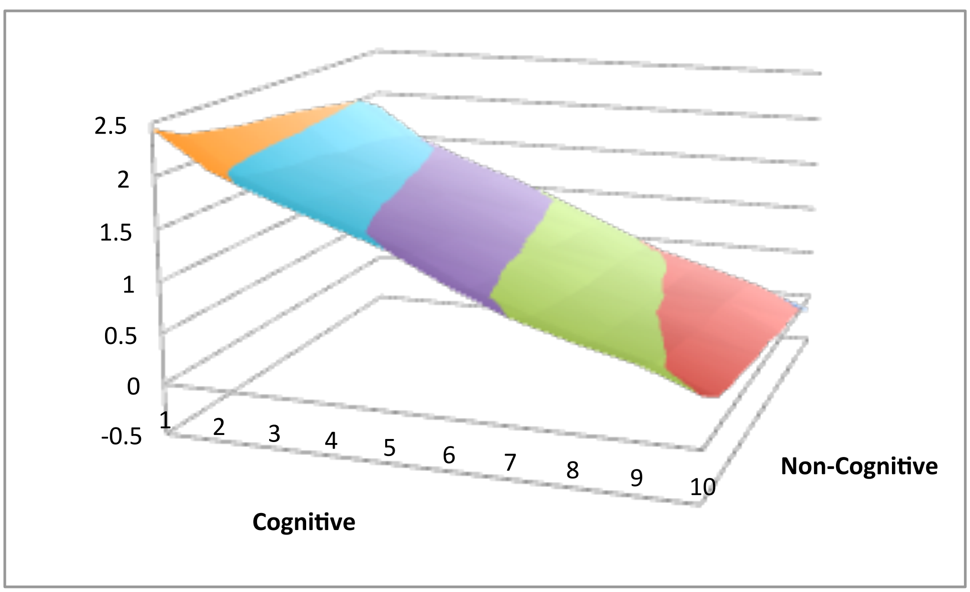
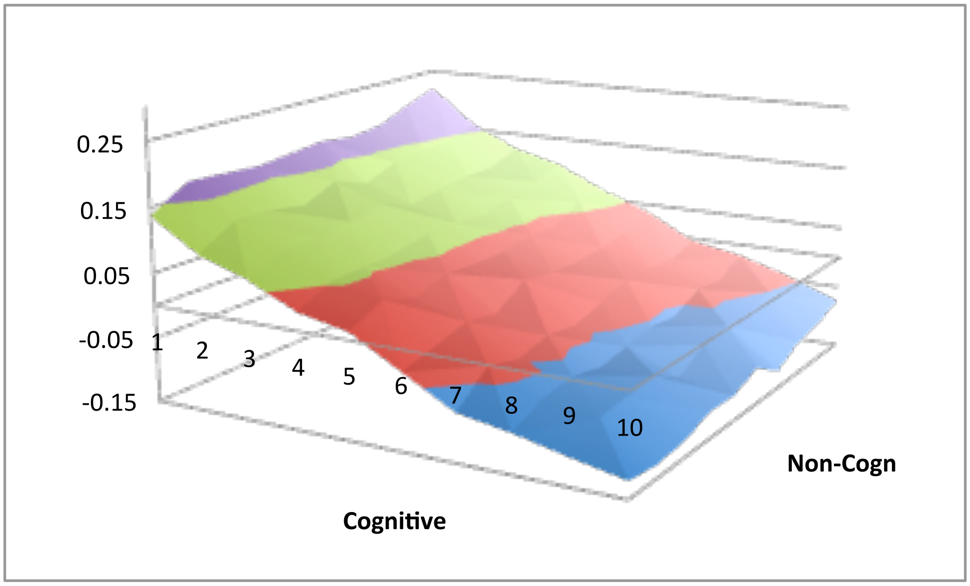
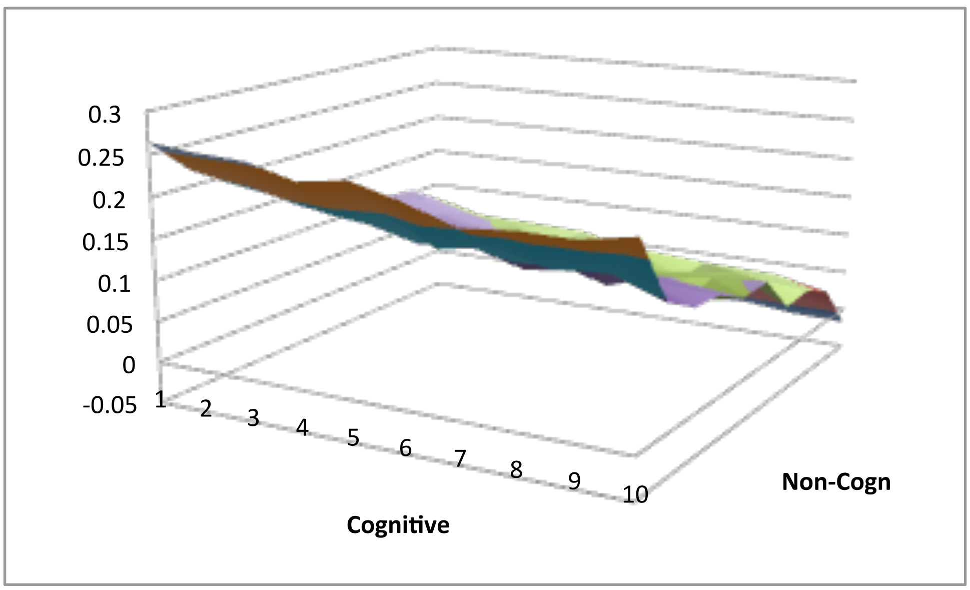


Fig. 1.3: ATE on Smoking

(a) Bullied



(b) Being a Bully



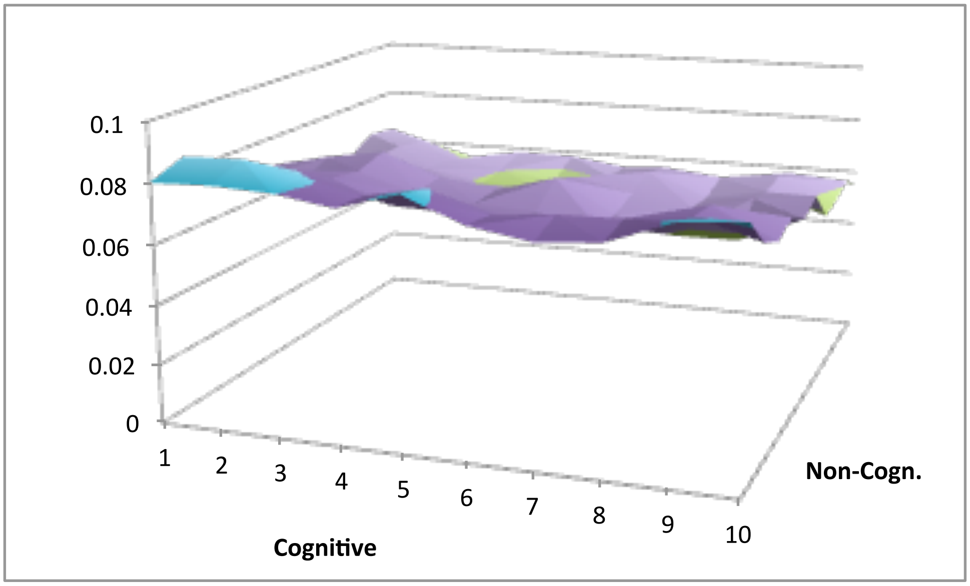
In subsection 1.6.2.2 we showed that former bullying victims are more likely to report poorer health. They reported feeling sick more frequently and were more likely to be diagnosed with mental health problems. Figure 1.4a shows that the treatment effect of being bullied on the likelihood of being sick is widespread and very stable across the entire cognitive-noncognitive space. The treatment effect is not as stable for the likelihood of having mental health issues (see Figures 1.5a and 1.5b). In fact, being bullied more than doubles the likelihood of having mental health issues for people with low cognitive skills and high non-cognitive skills. The relation is very different when the treatment considered is to be a bully. Although, never different from zero, the effect entirely depends on the level of non-cognitive skills.

Finally, Figures 1.6a and 1.6b show that the treatment effects of being a victim and being a bully on the likelihood of feeling stressed due to self-image is positive almost for the entire cognitive-noncognitive space. However these effects are zero, or close to zero, for people with high non-cognitive skills. That is bullying fosters insecurity, and it happens significantly more to those people with low non-cognitive skills.

All these results attest to the fact that skills not only affect bullying occurrence, but also, they mediate the extent to which these undesired behaviors affect subsequent outcomes, except those related to health.

Fig. 1.4: ATE on the Likelihood of Feeling Sick

(a) Bullied



(b) Being a Bully

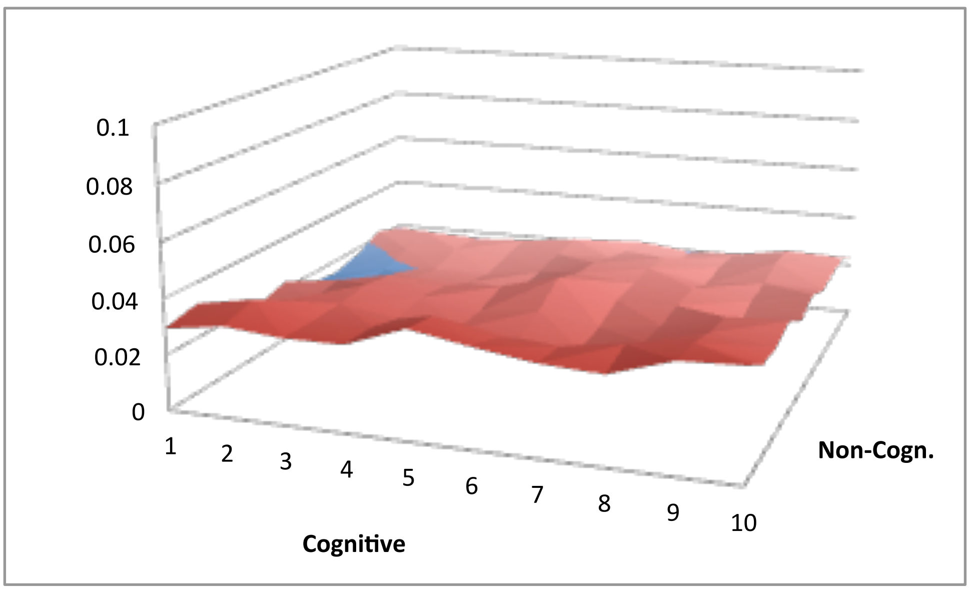
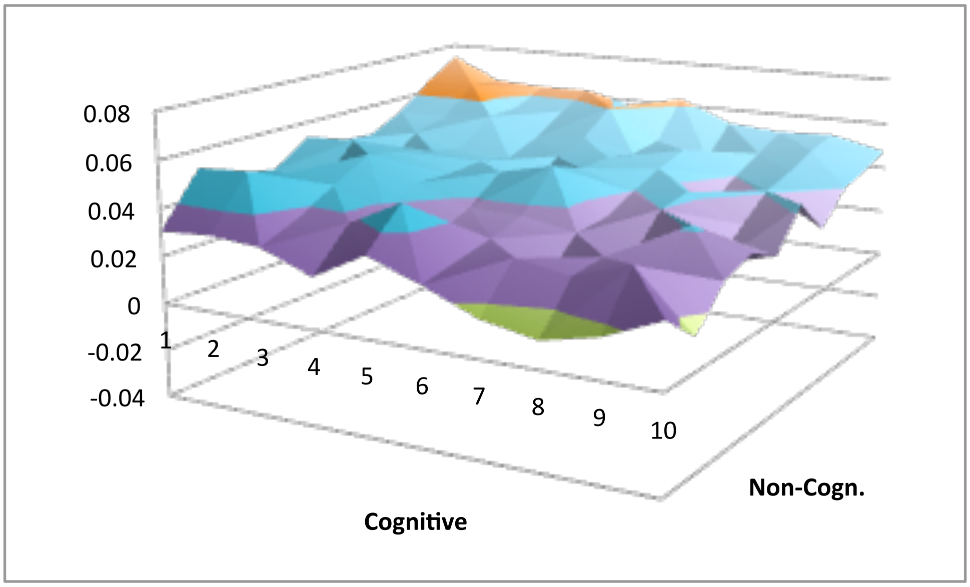


Fig. 1.5: ATE on Mental Health Problems

(a) Bullied



(b) Being a Bully

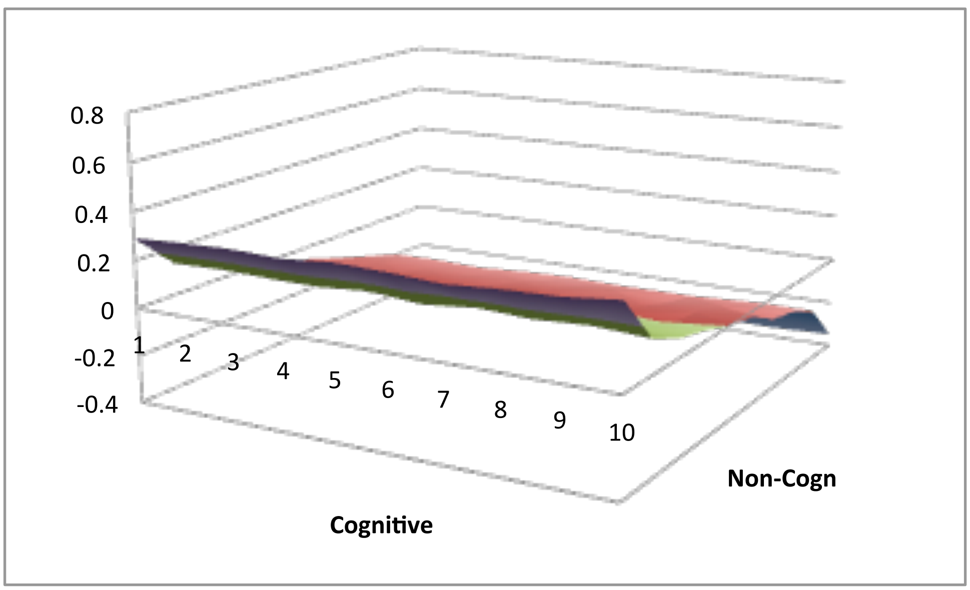
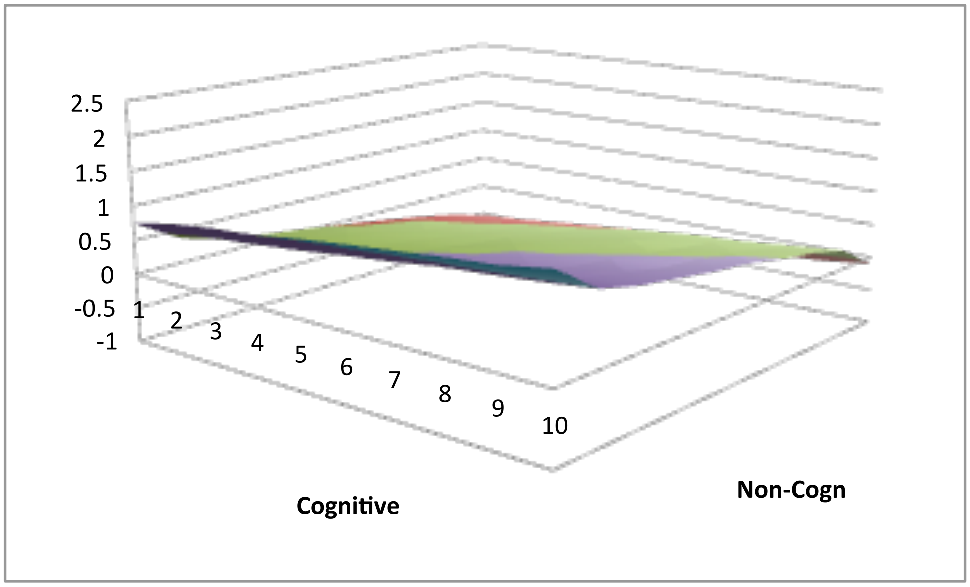
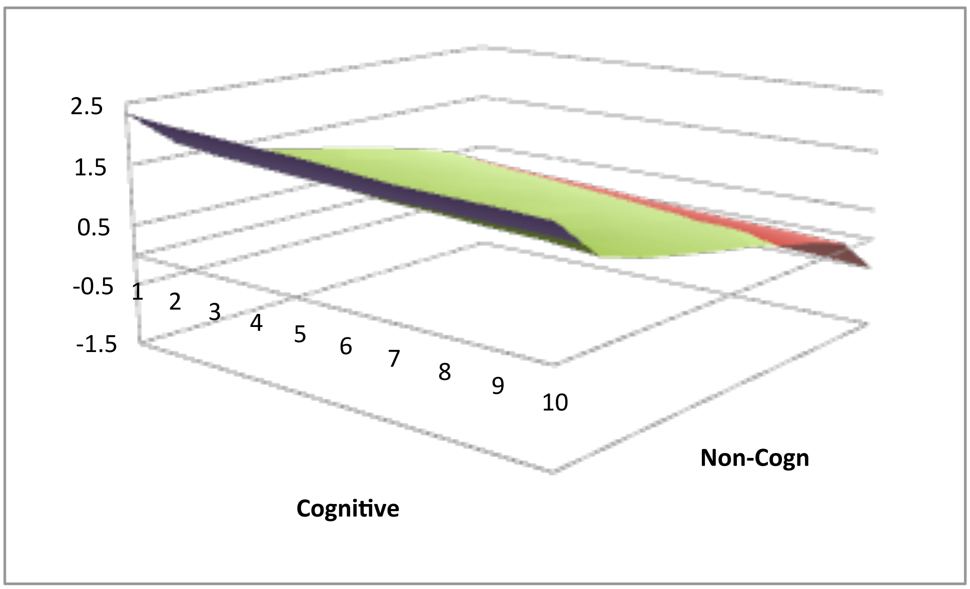


Fig. 1.6: ATE on Stress about Image

(a) Bullied



(b) Being a Bully



1.7 Bullying and Skill Investment

We have shown that non-cognitive skills are key determinants of bullying, both from the victim and from the perpetrator side. In an exploratory exercise we re-estimated the model from equation (1.18) to (1.22), but this time we included as controls variables that we believe can proxy skill investment. So if before we were controlling for θ_t^{NC}, now we will control for θ_t^{NC} and I_t^{NC}, where I_t^{NC} is a vector of non-cognitive skills' investment measures at time t. These investment measures include an index of parental control that measures whether the parents know where the youth is, who is he with and how long will he be there; an index of parental harmony that measures how much time the kid spends with their parents, whether she considers she is treated with affection by them, if she believes her parents treat each other well, and if her parents talk candidly and frequently with her; an index of parental abuse that measure whether the household is a violent setting. We also include the measures of school quality explained in Section (1.6.1.4).[14] If we are willing to assume assume that $\theta_t^{NC} + I_t^{NC} \approx \theta_{t+1}^{NC}$, we are then controlling for next period skills.

Our results are shown in the lower panel of Table 1.22. The top panel of the table reproduces the original results just for comparison. Our findings show that the introduction of investment controls reduces the point estimate of the effect of non-cognitive skills on the likelihood of being bullied, being a bully and being a cyberbully. Furthermore, less violence-prone parents and better schools reduce the incidence of bullying. In addition, higher parental control deters cyberbullying. We believe this is the case only for cyberbullying and not for physical bullying because while most of the physical bullying happens at school, cyberbullying happens outside school, probably at home or at friends' homes.

Hence, if we control for $t+1$ skills, we see that the part of it that comes from period t investment is a very strong bullying deterrent. Therefore, this exercise, exploratory at best, suggest that the inertia caused by low non-cognitive skills in previous periods

[14] Recall that the school quality measures are coded in a reverse scale where high numbers mean less school quality.

on higher likelihoods of being involved in bullying can be reversed through the modification of tangible scenarios like the improvement of schools –including teachers– and diminishing aggressive behavior within households.

Tab. 1.22: Adding Investment Controls

	Being Bullied		Bully		Cyberbully(New)	
VARIABLES	Coeff.	Std. err	Coeff.	Std. err	Coeff.	Std. err
No Inv. Contr						
Non-Cognitive	-0.074***	(0.015)	-0.025*	(0.013)	-0.066***	(0.020)
Cognitive	0.002	(0.007)	-0.007	(0.006)	-0.007	(0.009)
With Inv Contr						
Non-Cognitive	-0.054***	(0.016)	0.003	(0.014)	-0.046**	(0.021)
Cognitive	0.008	(0.007)	-0.002	(0.007)	-0.001	(0.010)
Inv Cont at t						
ParentControl	-0.001	(0.002)	-0.001	(0.002)	-0.006**	(0.003)
ParentHarmony	0.001	(0.002)	-0.002	(0.001)	0.001	(0.002)
ParentAbuse	0.006***	(0.002)	0.005**	(0.002)	0.006**	(0.003)
SchQuality	0.007***	(0.003)	0.008***	(0.002)	0.007**	(0.004)
SchEnviron	0.004***	(0.001)	0.004***	(0.001)	0.003*	(0.002)
Observations	2,862		2,862		2862	

*** p<0.01, ** p<0.05, * p<0.1

1.8 Conclusions

To the best of our knowledge this is the first attempt to quantify the effect of bullying on subsequent outcomes that explicitly deals with self-selection. We use a structural model that relies on the identification of latent cognitive and non-cognitive skill endowments to estimate the causal effect of being a victim of bullying and being a bully on outcomes like depression, smoking, health and stress. We find that non-cognitive skills reduce the likelihood of being a bully, a cyberbully and a victim of bullies. This is not the case for cognitive skills. We also showed that the model we estimate is able to recreate the observed choices and outcomes in our data. Therefore, we were confident we could simulate appropriate counterfactuals using our model. The simulation of the counterfactuals allowed us to calculate treatment effects in a self-selection setting.

Our findings indicate that bullying victims have higher incidence of depression, sickness, mental health issues and stress three years after being bullied. In the same vein, having been a bully increases the likelihood of depression, increases tobacco use and increases the feeling of stress. The sizes of these effects are by no means small, which attest to the fact that bullying is a heavy burden that needs to be carried for a long time. For instance, the incidence of health issues doubles due to bullying, and the chances of being a tobacco user doubles among former bullies.

Our findings indicate that the investment in skill development is key in any policy intended to fight bullying. We showed that the lack of skills increase the chances of being bullied and being a bully, we also showed that when using some skill investment controls, these effects went away. In addition, we showed that skills are mediators that can exacerbate or palliate the effect of bullying on later outcomes. Therefore, developing skills will not only reduce the incidence of bulling as there will be less people prone to be perpetrators and victims, but also the effects of these incidents would be lessened in a significant way.

2. THE DYNAMIC CONSEQUENCES OF BULLYING ON SKILL ACCUMULATION

2.1 Introduction

When Lim Jee-young, a South Korean woman from Daegu, read the suicide note left by her 13-year-old son, she found out that her young Seung-min had repeatedly been hit, robbed, burned and tortured by boys in his class.[1] Unfortunately, this horrible story is not unique. It repeats itself time and time again, not only in South Korea, but all around he world. The Columbine High School shooting in 1999 is still fresh in the collective memory of North Americans (US Department of Education, 2011),[2] as well as the story of Phoebe Prince, the 15-year-old girl that committed suicide in her house at South Hadley, Massachusetts, after having suffered several weeks of harassment and attacks while in school.[3] These and many other similar events—that exemplify the immense costs borne by bullying victims and communities in general—have increased the efforts made by societies to face and discuss bullying as behavioral issue among young people.[4]

Chapter 1 introduced the definition psychologists give to a bullying victim as a

[1] http://www.cnn.com/2012/07/25/world/asia/south-korea-school-bully/

[2] A US Secret Service report shows that 2/3 of 37 school shootings involved attackers that felt "persecuted, bullied, threatened, attacked or injured by others" (Vossekuil et al., 2002).

[3] http://www.nytimes.com/2010/03/30/us/30bully.html?ref=us. This event induced the Massachusetts Senate to pass an anti-bullying law that carries her name.

[4] Anti-bullying campaigns and laws have been implemented in the US, Canada, UK, Germany, Scandinavia, Colombia and South Korea.

person that is repeatedly and intentionally exposed to injury or discomfort by others in an environment where an imbalance of power exists (Olweus, 1997). The existence of an imbalance of power is a key feature of bullying because this creates a sense of defenselessness of the victim. He or she may be outnumbered, physically weaker or less psychologically strong than the bullies (Smith and Brain, 2000). As noticed in Chapter 1, Faris and Felmlee (2011) suggest that bullying thrives in contexts where individuals need to show peer group status. Not surprisingly, schools are the perfect setting for bullying.

Very little is known about the intermediate costs of bullying and its long-term consequences. This Chapter contributes in bridging this gap by exploring the two-way relation between bullying and cognitive and non-cognitive skills accumulation. *Cognitive skills* and *non-cognitive skills* are critical to the development of successful lives.[5] Hence, in this Chapter, I explore how school bullying hampers the development of successful adults by impeding optimal skill accumulation. In addition, I analyze how cognitive and non-cognitive skills are themselves important determinants of in-school victimization.

The proposed two-way relation between skills and bullying is based on facts about child victimization that have been widely established by the psychological literature. Psychologists have noticed not only that bullying victims suffer grave and long-lasting consequences in terms of their emotional well being (Olweus, 1997; Smith and Brain, 2000, among many others), but also that the likelihood of being a victim increases dramatically when the child has some behavioral vulnerability (Hodges et al., 1997).[6]

[5] See the definition of cognitive and non-cognitive skills in Chapter 1

[6] In fact, Hodges et al. (1997) show that the occurrence of chronic victimization needs two conditions. First, the victim needs to display a behavioral vulnerability that may not only make the kid irritant to his or her peer, but also may signal that he or she might not be able to defend him or herself as they "cry easily, are anxious, lack humor, lack self confidence and self esteem". Second,

To analyze this two-way relation, I develop a model that embeds peer-influenced skill contributions—like bullying—-in the skills accumulation process. I rely on the facts that skills are malleable and dynamic (Cunha and Heckman, 2007), and that many external conditions can influence the stock of skills an individual is endowed with at a given point in time (OECD, 2014). Little is known on how skill accumulation takes place, especially for non-cognitive skills. In particular, there is still a lot to be learned about the evolution of both cognitive and non-cognitive skills throughout childhood and teenage years, their dynamic interdependence, their two-way relationship with skill investment, the role social settings have on their development, and the long-term consequences of traumatic events and other skill-depleting forces like bullying. This Chapter intends to fill these gaps by quantifying the stock of skills lost to bullying by a given generation. The estimation strategy is based on the facts that skills beget skills (Cunha et al., 2010), that skills affect the occurrence of bullying (see Chapter 1), and that bullying affects skill accumulation. Bullying victimization depletes current skill levels and not only lowers skill accumulation, but also makes an individual more prone to experience bullying again in the future, creating a self-reinforcing mechanism that generates a big burden to be carried during adulthood. This intuition is developed in a dynamic model of skill formation in which the bullying event is treated as a negative shock that depletes the existing stock of skills changing negatively the skill accumulation path for the people involved.

Using a tractable model within the latent factor framework and extending the theoretical contributions of Cunha et al. (2010), I estimate the parameters that govern the process by which past skills determine future skills. I allow future skills to depend

victims occupy "a social position in the peer group that invites, disinhibits or permits aggressive attacks towards the child" .

on current skills, current investment choices and whether one was bullied. I also allow for the investment choices, which I treat as a latent factor as well, to depend on the level of skills. In addition, I allow the bullying event to depend on both the personal skills endowments and the skills distributions of the peer group each student is exposed to. This setting allows me to estimate both the *direct* (i.e., how bullying at t affects future skills at $t+1$) and *indirect* (i.e., how bullying at t affects future skills at $t+2$ through changes in skills at $t+1$) channels through which bullying affects skill accumulation.

This Chapter contributes to the economic literature in several ways. First, it inquires about the process of skill accumulation by providing a model that introduces peer characteristics into the dynamics of skill accumulation. Second, it analyzes the consequences of disruptive behavior in school in terms of skill depletion. Third, it extends my previous work on school bullying (see Chapter 1), where I found sizable consequences borne during adulthood, by providing additional insight into the channels through which high school bullying affects adult outcomes. Fourth, it allows the quantification of the long-run cost of bullying to a generation. That is, I can go beyond school absenteeism and in-school stress, and estimate skill endowments loses for life. In addition, this will open an auspicious research agenda on skill accumulation and victimization in schools. The analysis could be extended in many directions. For instance, we could inquire the extent to which bullying can affect skill accumulation of non-direct victims, just because of the disruptiveness of the event.

2.2 Related Literature

This work relates with two branches of the literature: bullying and skill formation. The related literature on bullying has already been reviewed in Section 1.2. However, let me add some additional references that are key in understanding different aspects of school bullying and it consequences in a dynamic setting. It is important to mention that just as he found that school and class size are not significant determinants of the likelihood of bullying, Olweus (1997) finds that nor are personal characteristics like disabilities, obesity, hygiene, posture and dress. However, he finds that victims are often smaller than attackers, and Lowenstein (1978) finds that victimized kids have more odd mannerisms than non-victimized kids. Kim (2005) shows that victimized children are more likely to suffer bed wetting, sleep difficulties, anxiety, loneliness and isolation, while Dake et al. (2003) find that bullied children are also more likely to feel lonely. All these analyses, although descriptive as they are unable to find causal effects, provide a critical input in the definition of the models I use in my own work.

In this Chapter, I contribute to the analysis of bullying literature by building on the findings of Chapter 1 and providing an explanation for how the difference in the outcomes observed in that Chapter materialize. That is, while Chapter 1 estimates the ATE of youth bullying on adult outcomes, in this present Chapter, I elucidate how these gaps are created, showing bullying as the triggering event that determines a divergence in skill accumulation paths.

Now let me turn to the literature on skill development, which remains scarce, even though it has benefitted from a number of recent contributions (Cunha et al., 2006, 2010). One of the main reasons for this scarcity is that it is difficult to directly measure cognitive and non-cognitive skills. At best, survey and administrative data contain measures that indirectly reflect underlying or latent skills. Structural models, like the one used in this dissertation, are needed to fully address this issue. Another difficulty arises due to the need of special longitudinal data sets that interview the subjects several times throughout particular periods of their life. Often this kind of data sets focus on labor, income and other economic dynamics and do not collect the measures needed to quantify skills.

What do we know so far? We know that skills are dynamic and malleable. That is, they depend on their past levels, they can be hindered and they can be fostered. Cunha et al. (2006) show that skills beget skills and therefore initial skill endowments and early accumulation are critical for the lifetime stock of skills. Based on this, they show that skill gaps between children from rich families and children from disadvantaged families start to widen at early ages. This gives foundation to the call for early childhood development and preschool interventions (Knudsen et al., 2006; Doyle et al., 2009). Skills beget skills not only though the natural process of getting the stock available at time t to $t+1$, but also though investment. There is some evidence on the fact that skills encourage skill investment. That is, skilled kids have higher levels of skill investment that lower skilled kids (Skinner and Belmont, 1993; Espinoza et al., 2014).[7]

The claim that skills are malleable is backed up by a series of papers that show that skill developing interventions were able to modify the stock of skills of the treated population. For instance, Heckman et al. (2010) show that the people treated by Perry Preschool Program have higher non-cognitive skills—although similar levels of cognitive skills—than the controls. The Socio-Emotional Learning programs have

[7] Although this feature of skill dynamics is widely proposed in Cunha et al. (2010), their theoretical claim of skills inducing higher levels of investment is only backed up by their empirical estimates in very early stages of life (i.e., before two years of age).

been widely reviewed as successful interventions that develop non-cognitive skills such as goal setting, conflict resolution and decision making (Payton et al., 2008). Cunha et al. (2010) show that skill developing interventions can compensate for low initial level of both cognitive and non-cognitive skills. There is evidence, however, that there are windows of opportunity outside of which skill malleability is lost (Knudsen et al., 2006). Cunha et al. (2006) argue that such window closes earlier for cognitive than for non-cognitive skills.

Besides the dynamism and malleability features of skills, recent literature has found that they strongly depend on different contexts the child grows in.[8] For instance, extensive literature finds that family background influences skill accumulation: children whose parents are more engaged in their upbringing are likely to have higher levels of both types of skill.[9] The quality of school inputs such as class size and teacher characteristics also affects non-cognitive skills (Fredriksson et al., 2013; Jackson, 2013). Skill endowments have even been found to depend on the level of stress a person was exposed to during childhood (McEwen and Seeman, 2006).

2.3 Skill Formation With Peer-Influenced Investment

Economic and psychological literature has found skills to be dynamic and prone to be influenced by many external conditions. Hence, my framework needs to incorporate five facts that result from this literature: i. skills beget skills, ii. skill development can be affected by investment choices, iii. past skills levels can affect next period skills indirectly by inducing skill investment, iv. bullying can hamper skill development,

[8] See OECD (2014) for a full framework about such contexts.

[9] See Hart and Risley (1995); Cunha et al. (2006); Heckman and Masterov (2007); Cabrera et al. (2007); Kiernan and Huerta (2008); Tough (2012)

and v. bullying depends also on the stock of cognitive and non-cognitive skills of each person and those of his or her peers. Therefore, I propose to augment the dynamic structure in Cunha et al. (2010) to explicitly incorporate these five facts. Let the stock of skills $S = \{A, B\}$ a person i that belongs to classroom c has at time $t+1$ (i.e., $\theta_{S,i\in c,t+1}$) be a result of a CES skill production function whose inputs are the stock of skills she had at time t ($\theta_{A,i\in c,t}$ and $\theta_{B,i\in c,t}$), the skill investment choices done between the two periods ($I_{S,i\in c,t}$), and the occurrence of a skill depleting shock like bullying ($M_{i\in c,t}$). Furthermore, I allow for the investment choice and the of bullying occurrence to be affected by the previous levels of skills.

$$\begin{gathered}
\theta_{S,i\in c,t+1} = \left[\gamma_{A,t}\theta^{\rho}_{A,i\in c,t} + \gamma_{B,t}\theta^{\rho}_{B,i\in c,t} + \gamma_{I,t}I^{\rho}_{S,i\in c,t+1} + \gamma_{M,t}M^{\rho}_{i\in c,t+1}\right]^{1/\rho} \\
I_{S,i\in c,t+1} = \alpha^{S}_{A,t}\theta_{A,i\in c,t} + \alpha^{S}_{B,t}\theta_{B,i\in c,t} + \nu_{t+1} \\
M_{i\in c,t+1} = \mathfrak{h}\left(\theta_{A,i\in c,t}, \theta_{B,i\in c,t}, \theta_{A,-i\in c,t}\right)
\end{gathered} \tag{2.1}$$

for $S = \{A, B\}$, where $\gamma_{M,t} = 1 - \gamma_{A,t} - \gamma_{B,t} - \gamma_{I,t}$ and $-i \in c$ indicates all individuals that belong to classroom c except i. Let N_c be the total number of students in classroom c, where $c \in C$.

This structure relates parental investment choices with victimization at $t+1$ indirectly through its effect on the stock of skills at t. This relies on the results of psychological research that indicates that responsive and supporting parenting practices are related with lower levels of bullying (Flouri and Buchanan, 2002). In particular, certain parental behaviors that hamper the development of locus of control on kids have been linked with in-school victimization (Ladd and Ladd, 1998).

In the system of equations (2.1), $\mathfrak{h}(\cdot)$ is a function that maps $\mathbb{R}^{N_c+1} \to \mathbb{R}$ representing a sufficient statistic that relates the skills available to other students in the classroom c with those of individual $i \in c$ that affects the likelihood of i being bullied. Through function $\mathfrak{h}(\cdot)$, I incorporate two stylized facts of bullying established by the psychological literature: i. that there are personal characteristics of the student that influences the chances of being bullied (i.e., behavioral issues), and ii. that there are characteristics of the peer group that set him or her apart from his or her classmates (e.g., lacks friends, is rejected by the peer-group) (Hodges et al., 1997). Function $\mathfrak{h}(\cdot)$ responds to the fact that bullying needs a social arena in which the imbalances of power take place allowing classmates to play different roles: victim, perpetrator and bystanders.[10] Therefore, the question that arises is: what separates bystanders from victims. Due to its social setting, one may be inclined to look for answers to this question in the social interactions literature as in Schelling (1971), Pollak (1976) and Manski (1993), where agents interact through their decisions. The problem with bullying is that no one *decides* to be a victim. Hence, while the social interactions literature explains "why do members of the same group tend to behave similarly" (Manski, 2000), we are instead interested in answering "why is this kid chosen among the rest". Hence, "selection into bullying" is non-random and, like in a *social interactions* situation, it depends on characteristics of the victim and its classmates, but in a very different way. The idea is that individual i with skills set $(\theta_{A,i,t}, \theta_{B,i,t})$ might be bullied in classroom c but not in classroom c'. This difference depends on the skill distribution of the other students that belong to each classroom. This defines a different dimension of peer-influenced consequences. Therefore, identification relies on the assumption that the allocation of individual i to classroom c was exogenous,

[10] Psychology literature has identified six types of classmates: ringleader bullies, follower bullies, reinforcers, defenders, bystanders and victims (Salmivalli et al., 1996). Due to data and computational restrictions, I compress the types of classmates to three: bullies, bystanders and victims.

and therefore the assignment of i's classmates is as good as random.[11]

The choice of the CES as the production function of skills responds to two main reasons. First, its capability to smoothly introduce investment and bullying as inputs together with past levels of skill. Second, it provides the curvature needed to explore complementarities between the inputs involved in the skill production function. In particular, the *static complementarity* ($\partial^2\theta_{S,t+1}/\partial I_{S,t+1}\partial\theta_{S,t}$) and the *dynamic complementarity* ($\partial^2\theta_{S,t+1}/\partial I_{S,t+1}\partial I_{S,t}$), concepts introduced by Cunha and Heckman (2008) to describe how the current stock of skills affect the productivity of skill investment, and how much of that productivity of investment is leveraged by past investment choices. I will use the same concepts to analyze the *skill depleting* power of the bullying event. Therefore the static complementarity on bullying will be given by

$$\frac{\partial^2\theta_{S,t+1}}{\partial M_{t+1}\partial\theta_{S,t}} \tag{2.2}$$

and the dynamic complementarity of bullying by

$$\frac{\partial^2\theta_{S,t+1}}{\partial M_{t+1}\partial M_t} \tag{2.3}$$

[11] This framework is particularly relevant in the South Korean context where after 1969 a "leveling policy" was introduced to regulate student placement. According to Kang (2007) "the law requires that elementary school graduates be randomly (by lottery) assigned to middle schools —either public or private— in the relevant residence-based school district." The leveling policy also makes the grouping of students by ability and achievement levels "extremely rare". These features let Kang (2007) claim that "the non-grouping (or ability mixing) in school exposes students to a classroom peer group that is nearly exogenously and randomly determined." Furthermore, the reader should note that unlike in the US, middle-school students in South Korea have a fixed classroom —and hence, classmates— for all subjects.

2.4 Empirical Strategy

As mentioned in Section 1.5, the key feature of the empirical strategy is the way it deals with the fact that underlying cognitive and non-cognitive skills and investment preferences are latent rather than observable.[12] They are not well defined entities with measurement scales and instruments, like height and weight are. Instead, these latent constructs need to be inferred from scores, called manifest variables, that can be directly observed and measured (Bartholomew et al., 2011).

In this Section, first I present how I use manifest scores to identify the latent variables of interest and then, based on that, I show how this allows me to estimate the full model.

2.4.1 Identification of Latent Factors' Distributions: From the Static to the Dynamic Setting

The core of the empirical strategy is the assumption of a linear relation between the manifest and the latent variables, that can be thought of as a production function of scores, whose inputs include both the individual observable characteristics and the latent endowments. In that sense, the empirical strategy incorporates the fact that the observed manifest values respond not only to the latent variables of interest (θ), but also to observable traits ($\mathbf{X}$) and random shocks (e^T) in the following form:

$$\mathbf{T}_t = \mathbf{X}_{t,T}\beta_t^T + \alpha_{\mathbf{t}}^{\mathbf{T},\mathbf{A}}\theta_t^A + {\alpha_{\mathbf{t}}}^{\mathbf{T},\mathbf{B}}\theta_t^B + \mathbf{e}_{\mathbf{t}}^{\mathbf{T}} \tag{2.4}$$

[12] In this dissertation I use the terms *latent variables* and *unobserved heterogeneity* interchangeably. While the term *latent variables* is widely used in statistics, the literature in labor economics prefers the term *unobserved heterogeneity* to differentiate it from the latent variable models that give the basis of probits, logits, censored and truncated estimations.

where $\mathbf{T}_t$ is a $L\times 1$ vector of measurements (e.g., test scores) at time t, $\mathbf{X}_{t,T}$ is a matrix with all observable controls for each measurement at time t and ${\alpha_{\mathbf{t}}}^{\mathbf{T,A}}$ and $\alpha_{\mathbf{t}}^{\mathbf{T,B}}$, are the loadings of the unobserved factors at time t. I assume that $\left({\theta_{\mathbf{t}}}^{\mathbf{A}}, {\theta_{\mathbf{t}}}^{\mathbf{B}}, \mathbf{X}_{t,T}\right) \perp \mathbf{e}_{\mathbf{t}}^{\mathbf{T}}$, that all the elements of the $L \times 1$ vector $\mathbf{e}_{\mathbf{t}}^{\mathbf{T}}$ are mutually independent and have associated distributions $f_{e_t^h}(\cdot)$ for every $h = 1, \ldots, L$.[13]

Carneiro et al. (2003), based on the insights of Kotlarski (1967), show that identification of the loadings in (2.4) (up to one normalization[14]) and the (diagonal) matrix of the variances of the latent factors Σ_θ needs three restrictions:

R1 Orthogonality of the factors (i.e., $\theta^A \perp \theta^B$).

R2 L to be at least $2k+1$, where k is the number of latent factors in the model.[15]

R3 The factor structure within the measurement system (2.4) needs to follow a triangular pattern like

$$\begin{bmatrix} \alpha^{T,A} & \alpha^{T,B} \end{bmatrix} = \begin{bmatrix} \alpha^{T_1,A} & \alpha^{T_1,B} \\ \alpha^{T_2,A} & \alpha^{T_2,B} \\ \alpha^{T_3,A} & \alpha^{T_3,B} \\ \alpha^{T_4,A} & \alpha^{T_4,B} \\ \alpha^{T_5,A} & \alpha^{T_5,B} \\ \alpha^{T_6,A} & \alpha^{T_6,B} \end{bmatrix} = \begin{bmatrix} \alpha^{T_1,A} & 0 \\ \alpha^{T_2,A} & 0 \\ 1 & 0 \\ \alpha^{T_4,A} & \alpha^{T_4,B} \\ \alpha^{T_5,A} & \alpha^{T_5,B} \\ \alpha^{T_6,A} & 1 \end{bmatrix} \tag{2.5}$$

which indicates that the first three manifest scores are affected by the first factor

[13] Identification of latent investment preferences in skill S, I_t^S, follows the same structure:

$$\mathcal{I}_{S,t} = \mathbf{X}_{t,\mathcal{I}}\beta_t^{\mathcal{I}_S} + \alpha_{\mathbf{t}}^{\mathcal{I}_\mathbf{S}} I_{A,t} + \mathbf{e}_{\mathbf{t}}^{\mathcal{I}_\mathbf{S}}$$

where $\mathcal{I}_{S,t}$ is a vector with the investment manifest scores.

[14] That is, one of the loadings $\alpha_t^{T,\cdot}$ of each factor should be set equal to 1, and the estimation of all the rest of the loadings should be interpreted as relative to those used as nummeraire.

[15] For simplicity, I will assume that I have $3k$ measurements in a measurement system like (2.4).

only, while the second three manifest scores are affected by both latent factors.[16]

However, these restrictions only apply to cases where there are no factor dynamics involved in the estimation. It is easy to see that Restriction R1 cannot be sustained if we believe there are dynamics governing the production of factor endowments at a given point in time. In particular, in a dynamic and intertwined process in which $\theta_{S,t+1} = g_S\left(\theta_t^A, \theta_t^B\right)$ for $S = \{A, B\}$, $\theta_{t+1}^A \not\perp \theta_{t+1}^B$ holds because of common past influences. That is, θ_{t+1}^A and θ_{t+1}^B are correlated because both share common inputs θ_t^A and θ_t^B, even if each latent factor has its own production function $g_A(\cdot,\cdot)$ and $g_B(\cdot,\cdot)$.

In order to get rid of the orthogonality assumption of contemporaneous latent factors R1 and still be able to identify the latent factors' distributions and loadings from a measurement system like (2.4), I need to rely on a factor structure different from the one required in Restriction R3. I need to assume a factor structure (2.5) where $\alpha^{T_6,A} = 0$ (i.e., for each latent factor, there is at least one manifest score that is only affected by that factor).

Theorem 1. *If $\theta_{t+1}^A \not\perp \theta_{t+1}^B$ and there is, at least one test score per latent factor that depends only on one latent factor, then the factor loadings of a measurement system like (2.4) are identified.*

[16] The loading structure of (2.4) depends entirely on the data available. Ideally, researchers have three measures for each factor, where each measure depends only on one factor. That is, in system (2.4) we will have the simplest version of (2.5):

$$\begin{bmatrix} \alpha^{T,A} & \alpha^{T,B} \end{bmatrix} = \begin{bmatrix} \alpha^{T_1,A} & \alpha^{T_1,B} \\ \alpha^{T_2,A} & \alpha^{T_2,B} \\ \alpha^{T_3,A} & \alpha^{T_3,B} \\ \alpha^{T_4,A} & \alpha^{T_4,B} \\ \alpha^{T_5,A} & \alpha^{T_5,B} \\ \alpha^{T_6,A} & \alpha^{T_6,B} \end{bmatrix} = \begin{bmatrix} \alpha^{T_1,A} & 0 \\ \alpha^{T_2,A} & 0 \\ 1 & 0 \\ 0 & \alpha^{T_4,B} \\ 0 & \alpha^{T_5,B} \\ 0 & 1 \end{bmatrix} \tag{2.6}$$

However, this is not often the case. There are many measures that depend on both latent factors. For instance, grades and education achievement scores may depend not only on a cognitive factor, but also on a non-cognitive one (Heckman et al., 2011a).

Proof. See Appendix K.1 □

Contrary to common practice latent factor literature, I do not impose normality to the distribution of the factors $f_{\theta^A,\theta^B}(\cdot,\cdot)$. Instead, I use the mixture of normals in order to achieve the flexibility required to mimic the true underlying distributions of the latent endowments. The mixture of normals not only grants flexibility in the type of distribution it is able to replicate, but also allows numerical integration using the Gauss-Hermite quadrature, which is particularly useful for calculating $E[f(X)]$ when $X \sim \mathcal{N}(\mu, \sigma^2)$ (Judd, 1998).[17] Therefore, the likelihood is

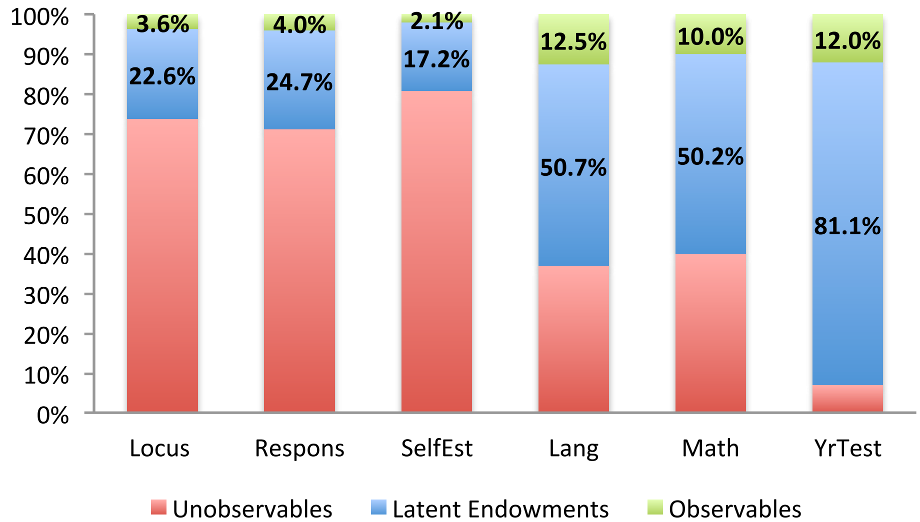
$$\mathcal{L} = \prod_{i=1}^{N} \int\int \Big[\ f_{e_t^1}\left(\mathbf{X}_{t,T_1}, T_{t,1}, \zeta^A, \zeta^B\right) \times \cdots \times f_{e_t^L}\left(\mathbf{X}_{t,T_L}, T_{t,L}, \zeta^A, \zeta^B\right) \ \Big] \Delta F_{\theta_t^A,\theta_t^B}\left(\zeta^A, \zeta^B\right) \tag{2.7}$$

where I integrate over the distributions of the factors due to their unobservable nature, obtaining $\hat{\beta}_t^T, \hat{\alpha}_t^{T,A}, \hat{\alpha}_t^{T,B}, \hat{F}_{\theta_t^A}(\cdot)$ and $\hat{F}_{\theta_t^B}(\cdot)$.

Using this latent variable framework allows me to use a construct that lacks metric and measuring instruments and disentangle the variation of interest (i.e., that one that comes from the unobserved heterogeneity) from the one generated by random shocks and the one that comes from exogenous observable traits like gender or age. In fact, variance decompositions of the test scores used for the estimation and presented in Figure 2.1 show that latent endowments explain between 5 to 9 times more the variation of the scores than the observable characteristics. However, Figure 2.1 also shows that even after controlling for latent endowments more that half of the variation of the scores is still unexplained. Hence, these findings are in line with the argument against the use test scores as proxies of abilities.

[17] The structural estimations presented in this Chapter were done using the `heterofactor` command for Stata developed by Miguel Sarzosa and Sergio Urzua (Sarzosa and Urzua, 2012).

Fig. 2.1: Decomposing Variances of Test Scores at $t = 1$



2.4.2 *Dynamic Estimation*

2.4.2.1 *Identification and Estimation Steps*

As explained above, the core of the empirical strategy is that $\theta_{S,t}$ and $I_{S,t}$ $\forall S, t$ are unobservable factors. Therefore, we need to estimate all the parameters of equations (2.4), and of the dynamic model (2.1) using maximum likelihood estimation procedures by integrating over the distributions that describe the latent factors. In this section, I will describe the steps and the identification sources involved in the estimation of the dynamic process described in (2.1).

Suppose the data we use follows individuals for two time periods: t and $t + 1$. Although in the actual estimations I use a triangular loading matrix like (2.5) with $\alpha^{T_6,A} = 0$, for simplicity suppose the manifest cognitive and non-cognitive scores have a loading structure presented in equation (2.6). Therefore the measurement system is the following:

$$\mathbf{T}_{A,t} = \mathbf{X}_{t,T}\beta_t^{T_A} + \alpha_{\mathbf{t}}^{\mathbf{T_A}}\theta_t^A + \mathbf{e}_{\mathbf{t}}^{\mathbf{T_A}} \tag{2.8}$$

$$\mathbf{T}_{B,t} = \mathbf{X}_{t,T}\beta_t^{T_B} + \alpha_{\mathbf{t}}^{\mathbf{T_B}}\theta_t^B + \mathbf{e}_{\mathbf{t}}^{\mathbf{T_B}} \tag{2.9}$$

$$\mathbf{T}_{A,t+1} = \mathbf{X}_{t+1,T}\beta_{t+1}^{T_A} + \alpha_{\mathbf{t+1}}^{\mathbf{T_A}}\theta_{t+1}^A + \mathbf{e}_{\mathbf{t+1}}^{\mathbf{T_A}} \tag{2.10}$$

$$\mathbf{T}_{B,t+1} = \mathbf{X}_{t+1,T}\beta_{t+1}^{T_B} + \alpha_{\mathbf{t+1}}^{\mathbf{T_B}}\theta_{t+1}^B + \mathbf{e}_{\mathbf{t+1}}^{\mathbf{T_B}} \tag{2.11}$$

$$\mathcal{I}_{A,t+1} = \mathbf{X}_{t+1,\mathcal{I}}\beta_{t+1}^{\mathcal{I}_A} + \alpha_{\mathbf{t+1}}^{\mathcal{I}_\mathbf{A}} I_{A,t+1} + \mathbf{e}_{\mathbf{t+1}}^{\mathcal{I}_\mathbf{A}} \tag{2.12}$$

$$\mathcal{I}_{B,t+1} = \mathbf{X}_{t+1,\mathcal{I}}\beta_{t+1}^{\mathcal{I}_B} + \alpha_{\mathbf{t+1}}^{\mathcal{I}_\mathbf{B}} I_{B,t+1} + \mathbf{e}_{\mathbf{t+1}}^{\mathcal{I}_\mathbf{B}} \tag{2.13}$$

$$M_{t+1} = \mathfrak{h}\left(\mathbf{X}_{t+1,M}\beta_{t+1}^M, \alpha_{\mathbf{t+1}}^{\mathbf{M}}\theta_{i\in c,t}, \alpha_{\mathbf{t+1}}^{\mathbf{M_c}}\theta_{-i\in c,t}\right) + e_{t+1}^M \tag{2.14}$$

where $\mathbf{T}_{S,\tau}$ is a 3×1 vector that contains each of the test scores associated to skill $S = \{A, B\}$ at time $\tau = \{t, t+1\}$, and $\mathcal{I}_{S,t+1}$ is a 3×1 vector that contains each of the investment measures made in skill $S = \{A, B\}$ at time $t + 1$. As shown in the previous section, we can use equations (2.8) and (2.9) to identify $\hat{F}_{\theta_{A,t},\theta_{B,t}}(\cdot,\cdot)$, and equations (2.12) and (2.13) to identify $\hat{F}_{I_{S,t+1}}(\cdot)$. Also, we can use (2.10) and (2.11) to consistently estimate $\hat{F}_{\theta_{A,t+1},\theta_{B,t+1}}(\cdot,\cdot)$ and $\hat{\beta}_{t+1}^{T_S}$. In consequence, I are able to construct the vectors

$$\xi_{A,t+1} = \mathbf{T}_{A,t+1} - \mathbf{X}_{t+1,T}\hat{\beta}_{t+1}^{T_A} = \alpha_{\mathbf{t+1}}^{\mathbf{T_A}}\theta_{t+1}^A + \mathbf{e}_{\mathbf{t+1}}^{\mathbf{T_A}} \tag{2.15}$$

$$\xi_{B,t+1} = \mathbf{T}_{B,t+1} - \mathbf{X}_{t+1,T}\hat{\beta}_{t+1}^{T_B} = \alpha_{\mathbf{t+1}}^{\mathbf{T_B}}\theta_{t+1}^B + \mathbf{e}_{\mathbf{t+1}}^{\mathbf{T_B}} \tag{2.16}$$

We now substitute the CES production function from (2.1) in the 3×1 measurement system for $\xi_{S,t+1}$. For instance, in the case when $S = A$ we have

$$\xi_{A,t+1} = \alpha_{\mathbf{t+1}}^{\mathbf{T_A}} \left[\gamma_{A,t}\theta_{A,i\in c,t}^{\rho} + \gamma_{B,t}\theta_{B,i\in c,t}^{\rho} + \gamma_{I,t}I_{S,i\in c,t+1}^{\rho} + \gamma_{M,t}M_{i\in c,t+1}^{\rho}\right]^{1/\rho} + \mathbf{e}_{\mathbf{t+1}}^{\mathbf{T_A}}$$

where $\gamma_{M,t} = 1 - \gamma_{A,t} - \gamma_{B,t} - \gamma_{I,t}$, from which I can identify $\gamma_{A,t}$, $\gamma_{B,t}$, $\gamma_{I,t}$ and ρ through a ML estimation assuming additive separability of the error term $\mathbf{e}_{\mathbf{t+1}}^{\mathbf{T_S}}$ and taking advantage of the fact that I have already identified $\alpha_{t+1}^{T_S}$. Hence, the likelihood function in the case where $S = A$ is:

$$\begin{aligned}
\mathcal{L} = \prod_{i=1}^{N} \int\int\int \; & f_{e_{t+1}^{T_{1,A}}}\left(\xi_{A,t+1}^{1} - \alpha_{t+1}^{T_{1,A}}\left[\begin{array}{c}\gamma_{A,t}\theta_{A,i\in c,t}^{\rho} + \gamma_{B,t}\theta_{B,i\in c,t}^{\rho} \\ +\gamma_{I,t}I_{A,i\in c,t+1}^{\rho} + \gamma_{M,t}M_{i\in c,t+1}^{\rho}\end{array}\right]^{1/\rho}\right) \\
& \times f_{e_{t+1}^{T_{2,A}}}\left(\xi_{A,t+1}^{2} - \alpha_{t+1}^{T_{2,A}}\left[\gamma_{A,t}\theta_{A,i\in c,t}^{\rho} + \gamma_{B,t}\theta_{B,i\in c,t}^{\rho} + \gamma_{I,t}I_{A,i\in c,t+1}^{\rho} + \gamma_{M,t}M_{i\in c,t+1}^{\rho}\right]^{1/\rho}\right) \\
& \times f_{e_{t+1}^{T_{3,A}}}\left(\xi_{A,t+1}^{3} - \left[\gamma_{A,t}\theta_{A,i\in c,t}^{\rho} + \gamma_{B,t}\theta_{B,i\in c,t}^{\rho} + \gamma_{I,t}I_{A,i\in c,t+1}^{\rho} + \gamma_{M,t}M_{i\in c,t+1}^{\rho}\right]^{1/\rho}\right) \\
& \times f_{\nu_{t+1}}\left(I_{A,i\in c,t+1}^{\rho} - \alpha_{A,t}^{A}\theta_{A,i\in c,t} - \alpha_{B,t}^{A}\theta_{B,i\in c,t}\right) \\
& \times f_{e_{t+1}^{M}}\left(M_{t+1} - \mathfrak{h}\left(\mathbf{X}_{t+1,M}\beta_{t+1}^{M}, \alpha_{\mathbf{t+1}}^{\mathbf{M}}\theta_{i\in c,t}, \alpha_{\mathbf{t+1}}^{\mathbf{M_c}}\theta_{-i\in c,t}\right)\right) \Delta F_{\theta_t^A,\theta_t^B}\left(\zeta^A,\zeta^B\right) dF_{I_{S,t+1}}\left(\zeta^I\right)
\end{aligned}$$

Note that given that I allow $\mathbf{X}_{t,T}$ and $\mathbf{X}_{t+1,T}$ to have a constant term, the estimated factors are centered at zero. That is, overall skill means are absorbed in the constant terms of the $\beta_{\tau}^{T_{3,S}}$ vectors for $\tau = \{t, t+1\}$ in equations (2.8) through (2.11). Estimation relies on that fact to identify ρ as shown in Theorem 2.

Theorem 2. *If $E[\theta_{A,t}] = 0$, $E[\theta_{B,t}] = 0$, the error e is additive separable in $\theta_{t+1} = f(\theta_t, \gamma, \rho) + e$, and $f(\theta_t, \gamma, \rho)$ is double differentiable , the parameter ρ is identified out of the fact that $E[\theta_{\cdot,t+1}] = 0$*

Proof. See Appendix K.2 □

Theorem 2 has an important implication regarding what I am able to capture in the dynamic process. The parameters estimated from (2.1) exclude effects due to the overall mean changes in skills. Suppose for instance, that everyone's cognitive skills increase on average by ϵ from t to $t+1$, such overall improvement does not affect the estimates of $\gamma_{A,t}$, $\gamma_{B,t}$, $\gamma_{I,t}$ and ρ . That is, my dynamic process will only capture idiosyncratic changes along the distributions. However, I can rely in the findings of Urzua (2008) who shows that under mild linearity assumptions in the measurement systems (2.8) through (2.11), I can claim that the mean of the skills is given by the coefficients in $\beta_{\tau}^{T_{3},S}$ associated with the constant term, call them $\beta_{\tau}^{T_{3},S}[1]$ for $\tau = \{t, t+1\}$. Therefore, I can retrieve the overall mean changes of skill S between t and $t+1$ (i.e., $\Delta_{t+1,t}^{S}$) by subtracting these constants. That is, $\Delta_{t+1,t}^{S} = \beta_{t+1}^{T_{3},S}[1] - \beta_{t}^{T_{3},S}[1]$.

2.4.2.2 *The Choice of $\mathfrak{h}(\cdot)$ and the Identification of the Bullying Equation*

As explained above, psychologists have established that in-school victimization occurrence requires some intrinsic characteristics of the person herself and characteristics of the person *vis-a-vis* the peer group (Hodges et al., 1997). The personal intrinsic traits are introduced in function $\mathfrak{h}(\cdot)$ through the observable and unobservable characteristics. The characteristics within the group, that is, the characteristics that "invite, disinhibits, or permits" attacks towards a given child are modeled by how uncommon in terms of particular traits the potential victim is. This uncommonness feature is important because it relates with several established facts in the psychological literature. First, that there needs to exist and imbalance of power (Olweus, 1997; Smith and Brain, 2000), therefore a kid with uncommon characteristics is less likely to have friends who can defend him or her (Bukowski et al., 1995). Also, bullies are more likely to attack those with no friends (Hodges and Perry, 1996). Furthermore, kids with uncommon characteristics are more easily regarded as weird and unlikeable, which fosters peer rejection (Hodges et al., 1997).[18]

The measure of rarity or uncommonness is materialized in my model by the number of classmates that lie inside an epsilon-ball in the skills or income space that is defined around those qualities for every kid. The intuition is that if your characteristics set you apart, meaning there are no kids similar to you, you have higher chances of being bullied. So, if $\nabla_{\psi,i\in c}(d)$ is the number of classmates of i that lie in an epsilon-ball with radius d in the space of characteristic ψ, then the $\mathfrak{h}(\cdot)$ function becomes

$$M_{i\in c,t+1} = \mathfrak{h}\left(\mathbf{X}_M, (\theta_A, \theta_B)_{i\in c,t}, \nabla_{\psi,i\in c}(d), \mu_c\right) \tag{2.17}$$

Where μ_c is a school fixed effect that responds to the fact that there are several school characteristics like teachers quality, or overall faculty tolerance to bullying that influence the likelihood of bullying victimization (Dake et al., 2003). In addition, I incorporate school district fixed-effects to properly take advantage of the school allocation randomness in the South Korean education system.

This way to introduce classmates' characteristics is also econometrically advantageous as it goes around the well known problem of peer-effect identification. According to Angrist (2014) randomness in peer allocation is not sufficient to identify peer-effects. He claims that, in order to prevent the unwanted existence of mechanical statistical forces that create spurious correlations, the econometrician needs that not

[18] Dake et al. (2003) show that students that scored higher on a scale of social acceptance were less likely to be bullied by their peers.

everyone within the group becomes affected or "treated" by the same peer-effect.[19] In my approach, the uncommonness measure allows for a different "treatment" for every observation to the point that, although everyone is affected by what happens inside their particular epsilon-ball, the relative position of those classmates that do not fall within that epsilon-ball is irrelevant.

In the estimations presented in this Chapter I consider bullying to be a dichotomous variable that takes the value of 1 if the person i was bullied and 0 if not. Hence, in this particular case, (2.17) becomes

$$E\left[M_{t+1}\right] = \Phi\left(\mathbf{X}_{M,i\in c}\beta^{M} + \alpha_{t+1}^{M_A}\theta_{A,i\in c,t} + \alpha_{t+1}^{M_B}\theta_{B,i\in c,t} + \alpha_{t+1}^{\nabla_{\theta_S}(d)}\nabla_{\psi,i\in c}\left(d\right) + \mu_c\right)$$

where $\Phi\left(\cdot\right)$ is the normal CDF.

2.5 Data

I empirically estimate the described model using the same data set used in Chapter 1. That is the Junior High School Panel (JHSP) of the Korean Youth Panel Survey (KYP). In this case, I choose to use these data motivated by two reasons. First, bullying is a very important social issue in the South Korean society, probably more so than anywhere else in the world as they have an active policy aimed to curve the incidence of suicide. Suicide figures in South Korea are striking. It causes 31.7 deaths per 100,000 people, the single highest rate in the world. Suicide is the largest cause of death for people between 15 and 24, killing 13 for every 100,000 people in this age range. One school-aged kid (10 to 19 years old) commits suicide each day.

[19] For instance, measuring peer-effects by introducing a classroom mean is invalid as everyone is being "treated" by that classroom mean which will create a tautological relation captured in the regression by the coefficient associated with the mean.

Suicide is often linked with school bullying. Statistics of the South Korean Education Ministry show that more than 77,000 students admitted to being bullied, and nearly 10 percent of those said they had considered suicide. In response to this, bullying has recently been placed at the center of social policy in South Korea by president Park Geun-hye as it has been considered to be one of the "four social evils" together with sexual assault, domestic violence and food contamination.[20] Since 2012, the government installed more than 100,000 closed-circuit cameras in school facilities to prevent bullying and prosecute its perpetrators.

The second reason to focus on South Korea is data availability and KYP-JHSP's unique sampling scheme that allows the identification of peer characteristics. The KYP-JHSP is a longitudinal survey that started in 2003 sampling full junior high-school classrooms (i.e., 14 year olds) from which all the students were interviewed.[21] They were interviewed once a year until 2008. Thus, they were followed though high-school and into the beginning of their adult life.

A more detailed description of the data set is presented in Section 1.3. Furthermore, Subsections 1.3.1 and 1.3.2 describe the procedures and data used to construct the non-cognitive and cognitive measures respectively.

2.5.1 Reported Bullying

Bullying, as all other personal characteristic that was collected in the KYP-JHSP, is self-reported by the students. It refers to events where they have been severely teased or bantered, threatened, collectively harassed, severely beaten, or robbed. Hence, even though psychologists have constructed a very wide definition for bullying which

[20] http://www.bbc.com/news/world-asia-26080052

[21] I find that there is at least one bully and one victim in every sampled classroom. This goes in line with the findings of Schuster (1999) in German schools.

I presented in the introduction of Chapter 1, the kids in the study respond to the most direct and less subtle versions of bullying. This is in line with the findings in several international studies (see Madsen, 1996; Smith and Levan, 1995; Smith et al., 1999, 2002) where children have been found to "focus on the more obvious and less subtle forms of bullying such as direct verbal and physical abuse and overlook indirect aggression" (Naylor et al., 2010).[22] In the same way, the reported incidence of bullying in the KYP-JHSP, presented in Table 2.1, is in line with the incidence reported in international studies (see Smith and Brain, 2000, for a summary).

Tab. 2.1: Incidence of Bullying by Wave

	Bullied	Bullied in $t = 1$	
Wave		No	Yes
1	.22499	·	·
2	.11198	.07377	.24271
3	.04768	.03270	.16871
4	.03428	.02275	.25517
5	.02231	.01941	.11340

2.6 Results

2.6.1 Skill Distributions

Tables 2.2 and 2.3 show the results of the estimation on the measurement system used to identify the joint skill distribution for $t = 1$ and $t = 2$, respectively. These distributions are presented in Figure 2.2a and 2.2b. They show that skill distributions are far from normal, and that there is a positive correlation between both dimensions of skills. In fact it is estimated to be of about 0.3869 for $t = 1$ and 0.358 for $t = 2$. This indicates that kids with high levels of one skill tend to have high levels of the

[22] This focus tends to collect versions of bulling that, to some extent, are more likely to happen among boys than among girls. Gender-specific analysis of bullying and its consequences is very important, but it will be addressed in later research.

other skills as well. An additional interesting feature of the joint skills distribution is the fact that the variance of non-cognitive skills increases for higher levels of cognitive-skills. Figure 2.3 shows kernel densities of non-cognitive skills for deciles 5, 6, 7 and 8 of the cognitive skills' distribution at $t = 1$. Hence, socio-emotional abilities, although positively correlated with cognitive skills, are less so for smarter kids.

Tab. 2.2: Identification of Skills at $t = 1$

VARIABLES	(1) Locus	(2) Irrespons	(3) Self-est	(4) Lang-SSc	(5) Math-Scie	(6) YearExam
Age (months)	-0.009*	0.015***	-0.014***	-0.008*	-0.009**	-0.014***
	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)
Male	0.150***	-0.055	0.170***	0.024	0.312***	-0.037
	(0.036)	(0.036)	(0.036)	(0.031)	(0.032)	(0.026)
Older Sibs	0.017	-0.006	-0.017	-0.045	0.032	0.009
	(0.033)	(0.033)	(0.034)	(0.029)	(0.030)	(0.026)
Young Sibs	0.016	-0.069**	0.023	0.077**	0.086***	0.086***
	(0.035)	(0.034)	(0.035)	(0.030)	(0.031)	(0.026)
lnInc_pc	0.074**	-0.104***	0.022	0.150***	0.145***	0.126***
	(0.033)	(0.033)	(0.033)	(0.029)	(0.029)	(0.025)
Urban	0.172***	-0.084	0.065	0.101**	0.067	-0.016
	(0.053)	(0.053)	(0.054)	(0.045)	(0.046)	(0.036)
Lives With:						
BothParents	0.209**	-0.313***	0.275***	0.314***	0.369***	0.235***
	(0.102)	(0.101)	(0.103)	(0.091)	(0.093)	(0.086)
OnlyMother	0.325**	-0.274**	0.428***	0.282**	0.354***	0.072
	(0.136)	(0.135)	(0.137)	(0.120)	(0.123)	(0.110)
Father Edu:						
2yrColl	0.133*	-0.147**	-0.022	0.145**	0.197***	0.205***
	(0.072)	(0.071)	(0.072)	(0.061)	(0.063)	(0.051)
4yrColl	0.141***	-0.142***	0.087**	0.317***	0.187***	0.245***
	(0.043)	(0.042)	(0.043)	(0.037)	(0.038)	(0.031)
GradSch	0.263***	-0.321***	0.148*	0.464***	0.288***	0.317***
	(0.076)	(0.075)	(0.076)	(0.063)	(0.065)	(0.049)
No-Cog Factor	1.213***	-1.258***	1	0.864***	0.934***	
	(0.090)	(0.096)	.	(0.110)	(0.115)	
Cogn Factor				0.533***	0.503***	1
				(0.026)	(0.027)	.
Constant	-0.755***	0.837***	-0.408**	-1.128***	-1.273***	-0.786***
	(0.174)	(0.173)	(0.175)	(0.157)	(0.160)	(0.148)
Observations	3,097					

Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Estimations include region fixed-effects. Factor distributions estimated using a mixture of two normals

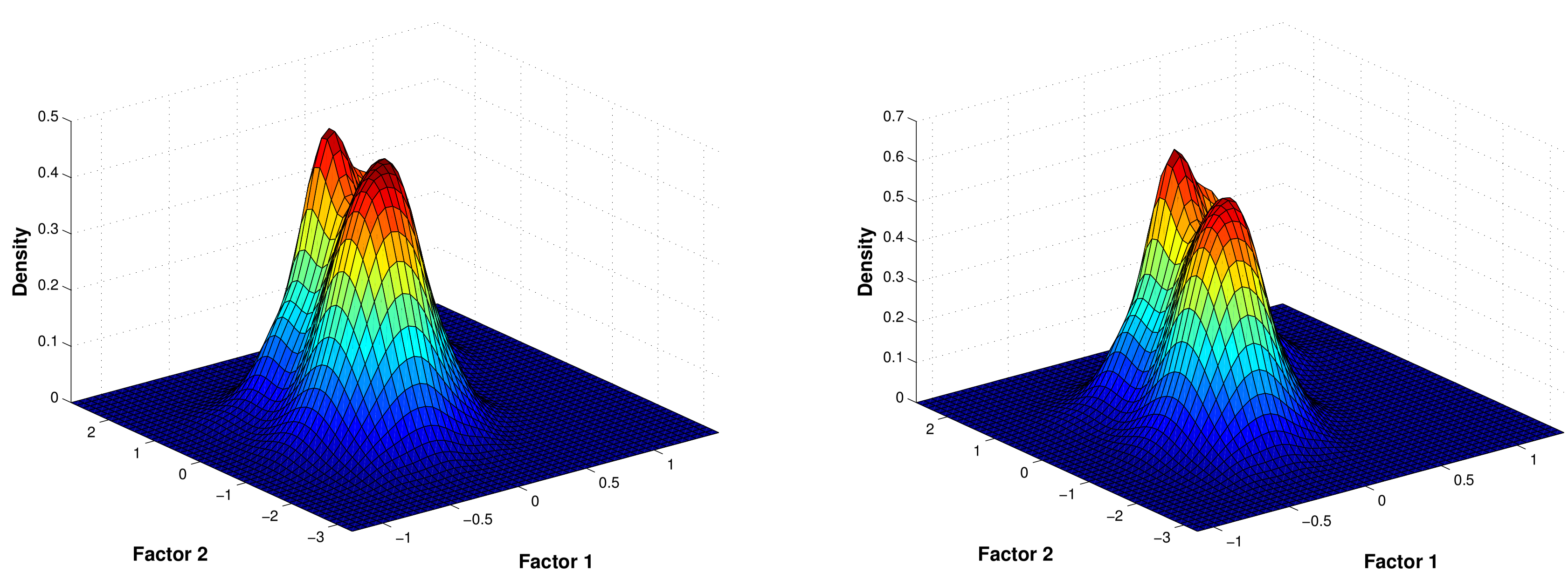
Tab. 2.3: Identification of Skills at 2

VARIABLES	(1) Locus	(2) Irrespons	(3) Self-est	(4) Lang-SSc	(5) Math-Scie	(6) YearExam
Age (months)	-0.023***	0.008	-0.008	-0.013***	-0.008*	-0.018***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)
Male	0.112***	-0.067*	0.136***	0.059*	0.366***	-0.058**
	(0.038)	(0.038)	(0.038)	(0.033)	(0.033)	(0.026)
Older Sibs	0.034	-0.016	0.055	-0.004	0.013	0.023
	(0.036)	(0.036)	(0.036)	(0.032)	(0.032)	(0.027)
Young Sibs	0.041	-0.092**	0.078**	0.143***	0.125***	0.082***
	(0.037)	(0.037)	(0.037)	(0.033)	(0.033)	(0.029)
lnInc_pc	0.087**	-0.039	0.068*	0.158***	0.164***	0.171***
	(0.037)	(0.037)	(0.037)	(0.034)	(0.033)	(0.030)
Urban	0.099*	-0.005	0.023	0.084*	0.059	-0.099**
	(0.058)	(0.058)	(0.058)	(0.050)	(0.050)	(0.039)
Lives With:						
BothParents	-0.079	-0.186**	0.062	0.286***	0.411***	0.302***
	(0.087)	(0.087)	(0.088)	(0.076)	(0.075)	(0.062)
OnlyMother	0.022	-0.244*	0.141	0.068	0.214*	0.220**
	(0.132)	(0.132)	(0.132)	(0.114)	(0.113)	(0.092)
Father Edu:						
2yrColl	-0.004	-0.209***	0.102	0.088	0.187***	0.178***
	(0.075)	(0.075)	(0.075)	(0.066)	(0.065)	(0.052)
4yrColl	0.112**	-0.166***	0.105**	0.295***	0.219***	0.253***
	(0.045)	(0.045)	(0.045)	(0.039)	(0.039)	(0.030)
GradSch	0.211**	-0.245***	0.119	0.358***	0.304***	0.346***
	(0.086)	(0.086)	(0.086)	(0.076)	(0.075)	(0.058)
No-Cog Factor	1.190***	-1.325***	1	1.351***	1.160***	
	(0.109)	(0.131)	.	(0.215)	(0.179)	
Cogn Factor				0.405***	0.461***	1
				(0.039)	(0.034)	.
Constant	-0.328	0.445**	-0.499**	-1.125***	-1.419***	-0.937***
	(0.200)	(0.201)	(0.201)	(0.181)	(0.180)	(0.163)
Observations	2,731					

Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Estimations include region fixed-effects. Factor distributions estimated using a mixture of two normals

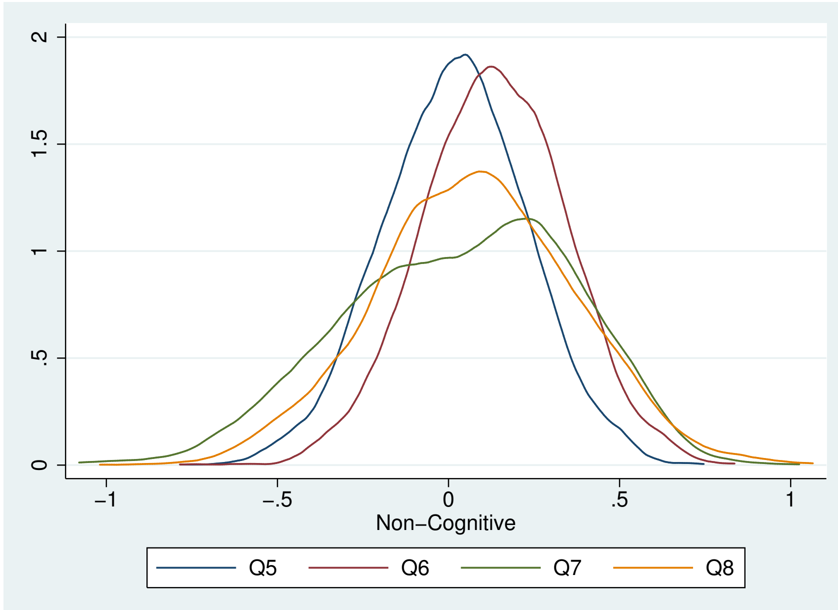
Fig. 2.2: Joint Skills Distribution

(a) $t = 1$ *(b)* $t = 2$



Note: Factor 1 refers to non-cognitive skills and Factor 2 refers to cognitive skills.

Fig. 2.3: Distribution of Non-cognitive by Decile of Cognitive Skills at $t = 1$



Note: Non-cognitive skills kernel densities for selected deciles of cognitive skills.

2.6.2 *Investment Factors*

As explained in Section 2.4, skills are not the only unobserved characteristics that enter the model. Investment choices made by the family are also unobserved factors. Hence, I estimate the underlying distributions from which the unobserved heterogeneity in investment comes from. That is, I estimate the measurement system described by equations (2.12) and (2.13).

The manifest variables used as measures of investment depend on the type of skill they intend to develop. I use measures of good parenting as indicator scores for investment choices in non-cognitive skills, namely parental physical and verbal abuse, parental control and parental harmony. The first measure indicates how often is the child beaten, physically hurt, yelled at or addressed in an inappropriate manner by the parents. Parental control relates to how well parents know where the kid is, who is she with, what is she doing and when is she coming back home. Parental harmony collects information related the level of care and interest in her life the kid feels from her parents.[23] The measures used to identify the cognitive skill investment factor relate to the enrollment in private tutoring of each kid. South Korean society is characterized by the high importance it gives to academic success. Hence, it is not uncommon for kids to be enrolled in after-school academic programs. By age 14, around four fifths of the kids in the survey attend some kind of tutoring. As manifest variables of cognitive skill investment I use a scale of how private—meaning how many classmates there are in every tutoring session, in a reversed scale—the tutoring is, the time spent in tutoring, and the cost of the tutoring.[24]

Tables 2.4 and 2.5 present the estimation of the investment in non-cognitive and cognitive skill factors respectively. The results show that the non-cognitive investment factor identified closely relates with good parental practices as it correlates positively with parental control and negatively with physical and verbal abuse. In the same way, the cognitive investment factor identified relates with the quality of after-class tutoring. It is positively correlated with how private the tutoring is, and how many hours the student spends in such after-class activities.

The non-cognitive investment distributions identified, presented in Figure 2.4, show a remarkable stability of the factor across wave. On the other hand, the cognitive investment distributions, presented in Figure 2.5, show two important characteristics. First, they are bimodal. That is the case because there are a proportion of kids that

[23] See Appendix I for a detailed explanation of the questions used to create each score

[24] The information used to inquire about investment in the development of cognitive skills related entirely on the usage of private tutoring. The first score, named type of tutoring, collect information of the nature of the extra-school classes taken. That is, whether the classes were entirely private, with few classmates, with many classmates, or through the internet. Students gave this type of information about their tutoring for every subject (e.g., language, math, science), and based on that I created aggregated measures. The second and third score used were straightforward: the amount of time and money spent in tutoring respectively.

Tab. 2.4: Identification of Unobserved Non-Cognitive Investment Factor

	$t = 1$			$t = 2$		
VARIABLES	Abuse	Control	Harmony	Abuse	Control	Harmony
Age (months)	0.0001	0.0005	-0.0081**	-0.0024	-0.0015	-0.0036
	(0.002)	(0.005)	(0.004)	(0.002)	(0.005)	(0.004)
Male	0.0434***	-0.2661***	-0.1638***	0.0169	-0.3143***	-0.2137***
	(0.016)	(0.033)	(0.026)	(0.014)	(0.034)	(0.028)
Older Siblings	-0.0019	-0.0153	0.0212	-0.0131	-0.0632*	-0.0025
	(0.015)	(0.031)	(0.022)	(0.013)	(0.033)	(0.028)
Young Siblings	-0.0079	0.0335	0.0186	-0.0017	-0.0136	0.0008
	(0.015)	(0.032)	(0.025)	(0.014)	(0.033)	(0.027)
lnInc_pc	-0.0217	0.0474	0.0899***	-0.0316**	0.1254***	0.1030***
	(0.015)	(0.033)	(0.026)	(0.014)	(0.033)	(0.027)
Urban	-0.0151	0.0256	0.0778**	-0.0277	0.1167**	0.1202***
	(0.024)	(0.050)	(0.037)	(0.021)	(0.051)	(0.044)
Lives: Both Parents	-0.1225***	0.1385*	0.1136*	-0.1082***	0.1474*	0.2351***
	(0.036)	(0.076)	(0.062)	(0.032)	(0.079)	(0.065)
Lives: Only Mother	-0.1391**	0.1400	0.2128**	-0.0861*	0.0581	0.3037***
	(0.054)	(0.113)	(0.086)	(0.045)	(0.109)	(0.087)
Father Edu: 2yColl	0.0555*	0.0407	0.1361***	0.0346	-0.0107	-0.0066
	(0.031)	(0.065)	(0.047)	(0.027)	(0.065)	(0.051)
Father Edu: 4yColl	-0.0287	0.0934**	0.0614**	-0.0411**	0.1372***	0.0998***
	(0.019)	(0.039)	(0.030)	(0.016)	(0.040)	(0.032)
Father Edu: GS	-0.1132***	0.3694***	0.1430**	-0.0750**	0.2273***	0.0854
	(0.034)	(0.072)	(0.059)	(0.030)	(0.072)	(0.055)
Non-Cogn Invest.	-0.1268***	0.5843***	1	-0.1269***	0.5564***	1
	(0.009)	(0.017)	.	(0.008)	(0.018)	.
Observations	2,988			2,968		

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Estimations include region fixed-effects. Factor distributions estimated using a mixture of two normals. Estimated intercepts not shown.

Tab. 2.5: Identification of Unobserved Cognitive Investment Factor

	$t = 1$			$t = 2$		
VARIABLES	Type Tutor	Tutor Time	Exp Tutor	Type Tutor	Tutor Time	Exp Tutor
Age (months)	0.0044	-0.0090*	0.0020	0.0056	-0.0101**	-0.0036
	(0.004)	(0.005)	(0.004)	(0.005)	(0.005)	(0.006)
Male	-0.0204	-0.0320	-0.0382	0.0999***	0.0830**	0.0962**
	(0.031)	(0.035)	(0.025)	(0.033)	(0.035)	(0.041)
Older Siblings	-0.0548*	-0.0016	-0.0329	-0.0619*	-0.0218	-0.0032
	(0.030)	(0.033)	(0.026)	(0.032)	(0.034)	(0.040)
Young Siblings	0.0043	0.0570*	0.0756***	0.0147	0.0502	0.1424***
	(0.030)	(0.034)	(0.024)	(0.033)	(0.035)	(0.042)
lnInc_pc	0.1166***	0.1539***	0.2429***	0.1102***	0.1233***	0.3401***
	(0.032)	(0.035)	(0.027)	(0.036)	(0.037)	(0.051)
Urban	-0.0846*	-0.1092**	-0.2407***	-0.2417***	-0.2372***	-0.5036***
	(0.048)	(0.053)	(0.039)	(0.051)	(0.054)	(0.062)
Lives: Both Parents	0.1304	-0.0170	-0.0108	0.2619***	0.2068**	0.2958**
	(0.094)	(0.103)	(0.085)	(0.095)	(0.098)	(0.134)
Lives: Only Mother	0.1159	-0.0154	0.0512	0.1271	0.1678	0.3272*
	(0.121)	(0.133)	(0.104)	(0.121)	(0.126)	(0.167)
Father Edu: 2yColl	-0.0247	0.1701**	0.1167**	-0.0264	0.0168	-0.0446
	(0.062)	(0.069)	(0.049)	(0.064)	(0.068)	(0.078)
Father Edu: 4yColl	0.0485	0.1247***	0.0477	0.0335	0.0912**	-0.0260
	(0.037)	(0.041)	(0.030)	(0.039)	(0.042)	(0.048)
Father Edu: GS	-0.0867	0.1108	0.0100	0.1455**	0.2851***	0.1395
	(0.067)	(0.075)	(0.051)	(0.073)	(0.076)	(0.096)
Cogn Investment	0.4747***	0.3025***	1	0.4185***	0.3387***	1
	(0.013)	(0.014)	.	(0.011)	(0.012)	.
Observations	2,918			2,761		

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Estimations include region fixed-effects. Factor distributions estimated using a mixture of two normals. Estimated intercepts not shown.

take no tutoring at all. Second, they are not stable in time. This responds to the fact that participation private tutoring falls as kids grow up.

Fig. 2.4: Unobserved Non-Cognitive Investment Factor

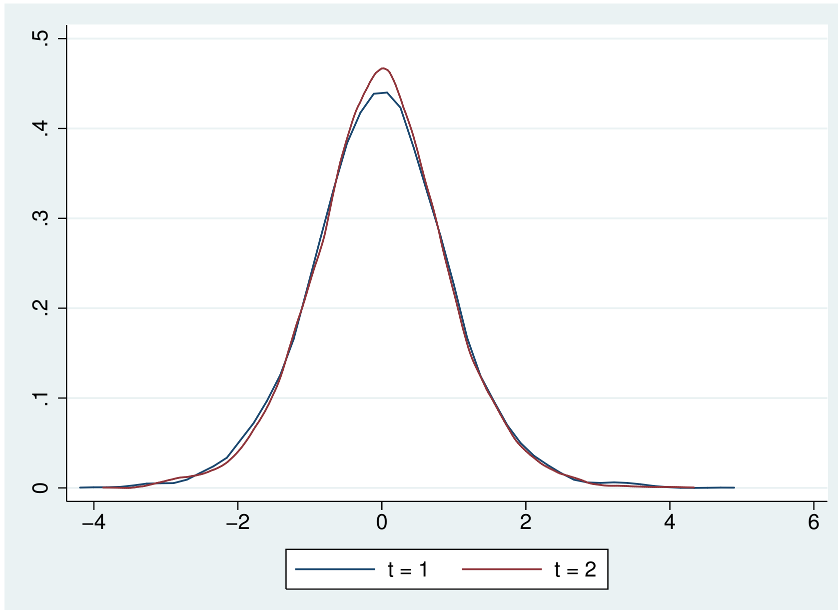
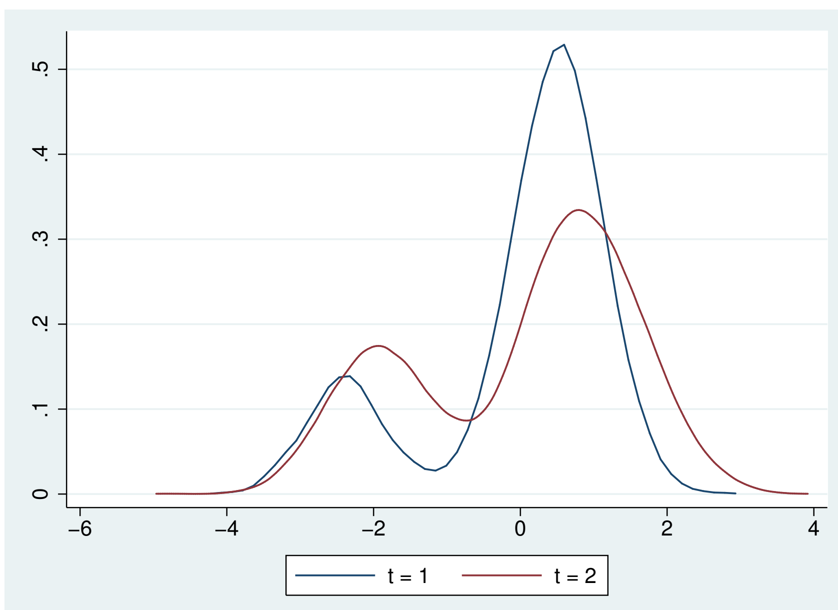


Fig. 2.5: Unobserved Cognitive Investment Factor



2.6.3 Results from the Dynamic Model

2.6.3.1 Incidence of victimization

Tables 2.6 and 2.7 show the relation between skills and selection into bullying.[25] In line with the results in Chapter 1, kids with less non-cognitive skills are significantly more likely to be bullied. Based on these estimates, the model allows me to quantify the probability of selection into bullying for each combination of skills at a given point in time. Figure 2.6a does this for $t = 1$ (age 14). It shows striking differences in the likelihood of being bullied depending on the level of non-cognitive skills. Kids in the first decile of non-cognitive skills have around a 50% chance of being bullied, while those chances for kids in the first decile is virtually zero. An standard deviation increase in non-cognitive skills reduces the likelihood of being bullied by 22 percentage points for the average student. That is, such increase in skills will reduce practically to zero the chances of being victimized.

Table 2.6 also shows the importance the relation between own and peer characteristics has in determining peer victimization. Controlling for observable characteristics and skill levels, youths who were placed in a school in which their non-cognitive skills are uncommon are significantly more likely to be bullied. The results indicate that the likelihood of victimization of the average student drops by one percentage point for each additional classmate with similar non-cognitive skill endowments he has. In-

[25] Given that the kids are already 14 years old by the first time they are interviewed, there is a possibility for the existence of joint causality between the contemporaneous measures of bullying and skills. I address this issue using the framework described in Hansen et al. (2004) and the exogenous variation that comes from the allocation of students to schools and classrooms. Hansen et al. (2004) require two additional assumptions for identification. First, the assumption of separability between the observed and unobserved part in every equation of the measurement system. Second, the assumption of orthogonality across the error terms in the complete measurement system. This last one is a very mild condition as every equation is being controlled not only for observable characteristics but also for the unobserved heterogeneity, which is theorized to be the only source of non-zero covariance between the unobservable parts of all the equations that comprise the full measurement system.

Tab. 2.6: Likelihood of Being Bullied at $t = 1$

	(1)		(2)		(3)		(4)		(5)	
VARIABLES					Coefficients					
Age (months)	0.0154*	(0.009)	0.0134	(0.008)	0.0152*	(0.009)	0.0149*	(0.009)	0.0148*	(0.009)
Male	0.3084***	(0.065)	0.2900***	(0.062)	0.3109***	(0.065)	0.3025***	(0.065)	0.3048***	(0.065)
Older Siblings	-0.0421	(0.058)	-0.0300	(0.055)	-0.0433	(0.058)	-0.0485	(0.058)	-0.0485	(0.058)
Young Siblings	-0.0878	(0.061)	-0.0721	(0.057)	-0.0888	(0.061)	-0.0922	(0.061)	-0.0926	(0.061)
lnInc_pc	-0.1005*	(0.058)	-0.0582	(0.054)	-0.1055*	(0.058)	-0.1221**	(0.059)	-0.1256**	(0.059)
Lives: Both Parents	-0.1960	(0.172)	-0.2848*	(0.158)	-0.1863	(0.172)	-0.1819	(0.172)	-0.1729	(0.172)
Lives: Only Mother	-0.1302	(0.228)	-0.1921	(0.212)	-0.1288	(0.228)	-0.1489	(0.228)	-0.1456	(0.229)
Father Edu: 2yColl	0.0411	(0.121)	0.0461	(0.117)	0.0398	(0.122)	0.0434	(0.121)	0.0430	(0.121)
Father Edu: 4yColl	-0.0577	(0.076)	0.0064	(0.072)	-0.0622	(0.076)	-0.0585	(0.076)	-0.0641	(0.076)
Father Edu: GS	0.1803	(0.134)	0.1666	(0.125)	0.1677	(0.134)	0.1673	(0.134)	0.1559	(0.134)
$E_{-i\in c}$[SchoolQty]	-0.0768	(0.112)			-0.0663	(0.112)	-0.0727	(0.111)	-0.0625	(0.111)
$E_{-i\in c}$[Non-Cog]			-0.2018	(0.345)						
$E_{-i\in c}$[Cognitiv]			0.5116	(0.495)						
Mass[Non-Cog]					-0.0401**	(0.019)			-0.0355*	(0.019)
Mass[Cognitive]					0.0257	(0.022)			0.0311	(0.022)
Mass[Income]							-0.0201*	(0.010)	-0.0194*	(0.011)
Non-Cogn	-1.9174***	(0.353)	-1.9029***	(0.333)	-1.9166***	(0.353)	-1.8980***	(0.350)	-1.8995***	(0.352)
Cognitive	0.2870***	(0.070)	0.2881***	(0.066)	0.2770***	(0.070)	0.2840***	(0.069)	0.2726***	(0.069)
					Marginal Effects at the Mean					
Mass[Non-Cog]					0.0115				0.0101	
Mass[Cognitive]					0.0064				0.0076	
Mass[Income]							-0.0057		-0.0055	
Non-Cogn	-0.5061		-0.5125		-0.5061		-0.5009		-0.5005	
Cognitive	0.0757		0.0776		0.0731		0.0749		0.0718	
Observations	2,805		3,097		2,805		2,805		2,805	

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Estimations include region fixed-effects. Estimations include school fixed-effects. Mass[] refers to the number of observations within a window of 10% of a SD around observation i. The marginal effect of the Mass[] variables are calculated based on the discrete change in the number of people inside the window from 0 to 1. Coefficients associated to the intercepts and the urban control not shown.

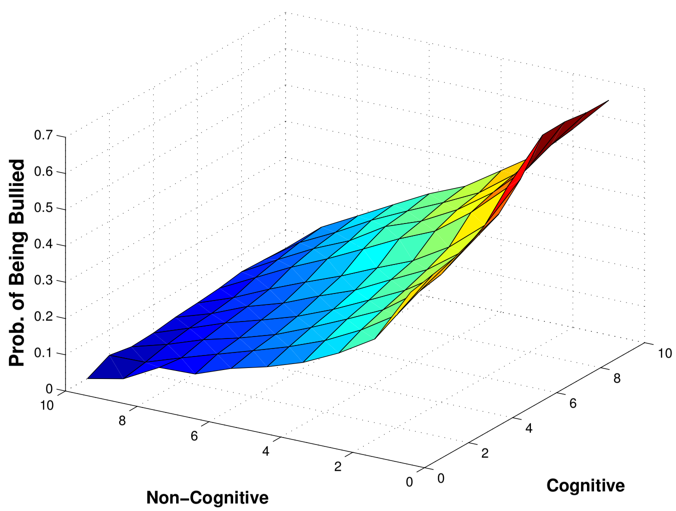
Tab. 2.7: Likelihood of Being Bullied at $t = 2$

VARIABLES	(1)		(2)		(3) Coefficients		(4)		(5)	
Age (months)	0.0077	(0.011)	0.0072	(0.011)	0.0069	(0.011)	0.0071	(0.011)	0.0068	(0.011)
Male	0.3381***	(0.083)	0.2878***	(0.082)	0.3440***	(0.084)	0.3340***	(0.083)	0.3442***	(0.084)
Older Siblings	-0.0233	(0.075)	-0.0643	(0.073)	-0.0174	(0.075)	-0.0217	(0.074)	-0.0176	(0.075)
Young Siblings	-0.1520*	(0.081)	-0.1851**	(0.081)	-0.1437*	(0.081)	-0.1504*	(0.080)	-0.1440*	(0.081)
lnInc_pc	0.0249	(0.078)	-0.0090	(0.076)	0.0182	(0.078)	0.0074	(0.078)	0.0184	(0.078)
Lives: Both Parents	0.0305	(0.188)	0.0694	(0.184)	0.0266	(0.189)	0.0509	(0.188)	0.0277	(0.189)
Lives: Only Mother	0.3882	(0.257)	0.3485	(0.255)	0.3941	(0.259)	0.3764	(0.256)	0.3939	(0.259)
Father Edu: 2yColl	0.0347	(0.151)	-0.0006	(0.152)	0.0292	(0.152)	0.0273	(0.151)	0.0291	(0.152)
Father Edu: 4yColl	-0.1319	(0.098)	-0.1199	(0.096)	-0.1419	(0.099)	-0.1335	(0.098)	-0.1417	(0.099)
Father Edu: GS	0.1202	(0.172)	0.1072	(0.169)	0.1116	(0.173)	0.1148	(0.171)	0.1116	(0.173)
$E_{-i\in c}$[SchoolQty]	0.2132	(0.184)			0.2327	(0.184)	0.2168	(0.183)	0.2370	(0.184)
$E_{-i\in c}$[Non-Cog]			-0.4134	(0.548)						
$E_{-i\in c}$[Cognitiv]			0.9342*	(0.536)						
Mass[Non-Cog]					-0.1118**	(0.056)			-0.1118**	(0.056)
Mass[Cognitive]					-0.0397	(0.056)			-0.0392	(0.057)
Mass[Income]							-0.0195	(0.014)	-0.0014	(0.015)
Non-Cogn	-2.3146***	(0.588)	-2.4713***	(0.614)	-2.3480***	(0.592)	-2.2924***	(0.586)	-2.3471***	(0.592)
Cognitive	0.3436***	(0.093)	0.3488***	(0.094)	0.3612***	(0.094)	0.3402***	(0.093)	0.3608***	(0.095)
					Marginal Effects at the Mean					
Mass[Non-Cog]					-0.0178				-0.0177	
Mass[Cognitive]					-0.0056				-0.0055	
Mass[Income]							-0.0030		-0.0002	
Non-Cogn	-0.2922		-0.3269		-0.3152		-0.3176		-0.3146	
Cognitive	0.0463		0.0460		0.0485		0.0471		0.0484	
Observations	2,553		2,725		2,550		2,553		2,550	

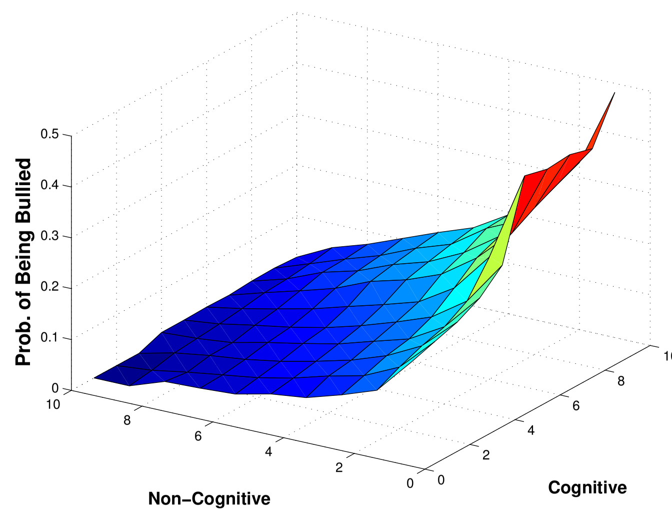
Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Estimations include region fixed-effects. Estimations include school fixed-effects. Mass[] refers to the number of observations within a window of 30% of a SD around observation i. The marginal effect of the Mass[] variables are calculated based on the discrete change in the number of people inside the window from 0 to 1. Coefficients associated to the intercepts and the urban control not shown.

Fig. 2.6: Probability of Being Bullied

(a) $t = 1$



(b) $t = 2$



terestingly, the same occurs in terms of income. Bullying probability falls by half a percentage point for each additional classmate that has a family income level that falls within the epsilon-ball defined around the family income level of the prospective victim.

As I showed in Table 2.1, the incidence of bullying falls as the kids become older. In fact from wave one to wave two the incidence of being bullied is cut by more than half, falling from 22.5% to 11.12%. However, the results in Table 2.7 and in Figure 2.6b show that victimization is not reduced across the board. In fact, while the probability of being bullied has fallen virtually to zero for the kids in the last seven non-cognitive skill deciles, the kids in the first decile remain with similar probabilities of being bullied than a year before, around four times the average. Hence, it is fair to say that in time bullying becomes more selective of its victims, overwhelmingly focusing on kid that lack non-cognitive abilities.

Table 2.7 indicates that the relation between own and peer characteristics becomes even more important at age 15. Non-cognitive skills uncommonness significantly increases victimization likelihood even more so than before. My results indicate that bullying probability falls by 1.7 percentage points per each additional classmate whose non-cognitive skills fall within the epsilon-ball defined around the non-cognitive skills of the prospective victim.

An important distinction with the empirical strategy in Chapter 1 is that in the present Chapter I allow for correlated skills. Once I allow for this correlation to exist, I find a positive effect of cognitive skills on the likelihood of being bullied; an effect that is not reported in Chapter 1. This is a very interesting finding as it reflects the fact that the kids that are more likely to be bullied are those who are smart but lack non-cognitive skills. However, Figures 2.6a and 2.6b show that any positive effect cognitive skills may have is dwarfed by the negative effect of non-cognitive skills.

2.6.3.2 Skills Production

Table 2.8 presents the results of estimating the system described by (2.1). In the first part of the table, I present the raw estimates of the dynamic parameters. Using these structural parameters—including the one related to selection into bullying, I am able to fully recreate the dynamic process. This process is presented in Figures 2.7a, 2.7c, 2.8a and 2.8c. Figures 2.7a and 2.8a show that high non-cognitive skills produce high future non-cognitive skills, and that marginal increments of those initial skills are very productive (i.e., non-cognitive skills self-productivity ${}^{\partial\theta_{t+1}^{NC}}/_{\partial\theta_t^{NC}} > 0$ for the entire $\left(\theta_t^{NC}, \theta_t^{C}\right)$ space, see Figure 2.9a). These figures also demonstrate that cognitive skills are unimportant in non-cognitive skill production process except for the fact that higher initial cognitive skills make the marginal increments of the initial non-cognitive skills more productive (i.e., ${}^{\partial^2\theta_{t+1}^{NC}}/_{\partial\theta_t^{NC}\partial\theta_t^{C}} > 0$).

Figures 2.7c and 2.8c show that the production of cognitive skills relies heavily on past levels of cognitive skills. Although the existing levels of non-cognitive skills contribute in the production process of cognitive skills, their contribution is small compared to that of the existing stock of cognitive skills. For instance, going from decile 1 to decile 10 in non-cognitive skills distribution has the same effect on the production of cognitive skills as increasing the cognitive skills input by one decile. Figure 2.9c shows that, at age 14 (i.e., $t = 1$), the self-productivity of cognitive skills is higher among non-cognitive skilled people. Analogously, Figure 2.9d shows that the productivity of non-cognitive skills in producing next period cognitive skills is higher among kids with high initial cognitive skills.

My results indicate that there is a strong path dependence in which skills produce skills, setting a high cost in terms of future stock of skills for those who start the accumulation process in the lower quantiles of the skill distribution. My results also show that this path dependence is not reversed by investment choices. In fact, Table 2.8 and Figure 2.11 show that investment choices in non-cognitive skills depend greatly on the past level of non-cognitive skills, and investment choices in cognitive skills depend greatly on past levels of that skill in the first place. Hence, people with high skills not only pass their high stock on to the next period, but also they are more prone to invest in their development.

2.6.3.3 Effects of Bullying on Skill Production and Future Bullying

The right panel of Table 2.8 shows the effect of bullying on the accumulation of cognitive and non-cognitive skills. To calculate this, I compare the next period skills of those who would be selected into bullying with those who would not, evaluated at the skills' mean. That is, $E\left[\hat{\theta}^{S}_{t+1}|\bar{\theta}^{NC}_{t},\bar{\theta}^{N}_{t},M_{t+1}=1\right]-E\left[\hat{\theta}^{S}_{t+1}|\bar{\theta}^{NC}_{t},\bar{\theta}^{N}_{t},M_{t+1}=0\right]$ for $S=\{NC,C\}$. For age 14, I find there is a statistically significant effect of -0.0631 of bullying on next period non-cognitive skill accumulation. In order to put into perspective this figure, note that the standard deviation of θ^{NC}_{t+1} is 0.4461. Therefore, my results indicate that bulling reduces non-cognitive skill accumulation by 14.15% of a standard deviation for the average kid. This is a very sizable effect. It implies a reduction of 19.11% of a standard deviation in the language test score, and a reduction of 16.41% of a standard deviation in the math test score. The differential effect of bulling in next period non-cognitive skills depending on previous skills levels

Tab. 2.8: Dynamic Estimation $\theta^S_{t+1} = f\left(\theta^{NC}_t, \theta^C_t, I^S, Bullied\right)$

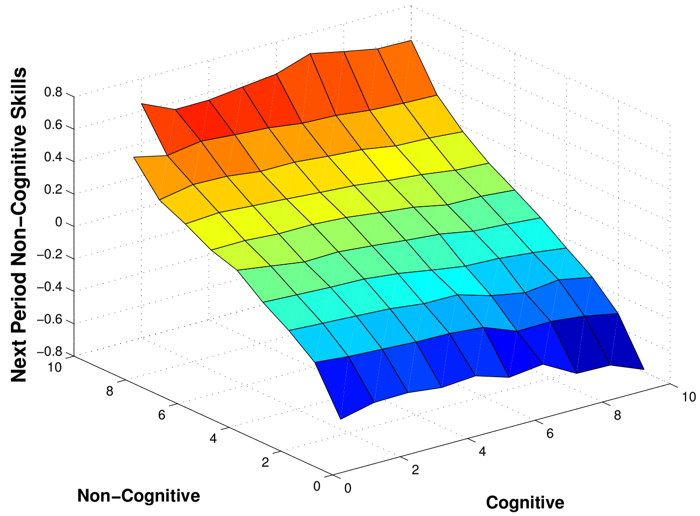
	$(t=1)$ θ^{NC}_{t+1}	$(t=1)$ θ^{C}_{t+1}	$(t=2)$ θ^{NC}_{t+1}	$(t=2)$ θ^{C}_{t+1}
θ^{NC}_t	1.0513***	0.3505***	1.4077***	0.3802***
	(0.039)	(0.095)	(0.067)	(0.020)
θ^{C}_t	-0.0569***	0.7367***	-0.1329***	0.8239***
	(0.015)	(0.026)	(0.019)	(0.007)
I^{NC}_{t+1}	0.0570**	0.0525***	0.0447	-0.0166***
	(0.024)	(0.009)	(0.029)	(0.003)
M_{t+1}	-0.0514**	-0.1298*	-0.3195***	-0.1873***
	(0.025)	(0.743)	(0.042)	(0.015)
ρ	0.3755	-0.1431	0.2350*	0.6774***
	(0.385)	(0.1217)	(0.127)	(0.057)
$\alpha^{NC}_{NC,t}$	0.7917***	0.1712	1.1452***	0.5654***
	(0.122)	(0.190)	(0.158)	(0.1955)
$\alpha^{NC}_{C,t}$	-0.033	0.2519***	-0.0421	0.3578***
	(0.031)	(0.046)	(0.036)	(0.052)
Obs.	2,345		2,233	

$E\left[\theta^{NC}_{t+1} \mid \bar{\theta}^{NC}_t, \bar{\theta}^{N}_t, M_{t+1}=1\right] - E\left[\theta^{NC}_{t+1} \mid \bar{\theta}^{NC}_t, \bar{\theta}^{N}_t, M_{t+1}=0\right]$	$(t=1)$	$(t=2)$
θ^{NC}_{t+1}		
Raw Effect	-0.0631**	-0.3763***
	(0.029)	(0.065)
As StdDev of θ^{NC}_{t+1}	-0.1415	-0.7280
θ^{C}_{t+1}		
Raw Effect	-0.1199	-0.2957***
	(0.163)	(0.022)
As StdDev of θ^{C}_{t+1}	-0.1566	-0.3504
Obs.	2,345	2,233

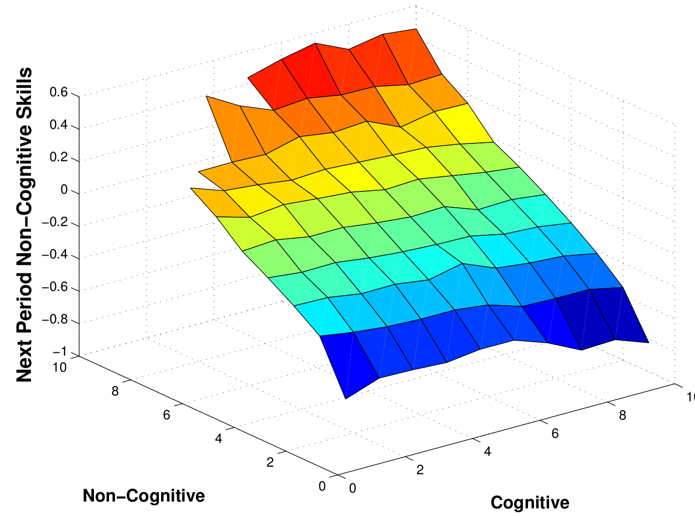
Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Fig. 2.7: θ_{t+1}^{S} as a function of θ_t^{NC} and θ_t^{C} $(t = 1)$

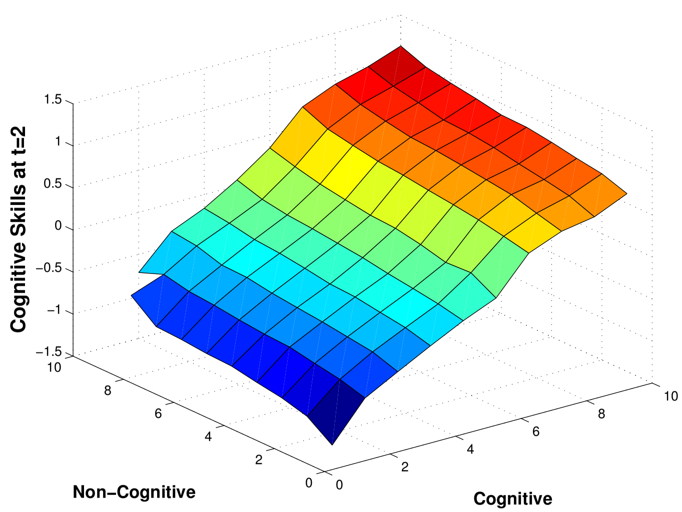
(a) Non-Cognitive: Everyone



(b) Non-Cognitive: Bullying Victims



(c) Cognitive: Everyone



(d) Cognitive: Bullying Victims

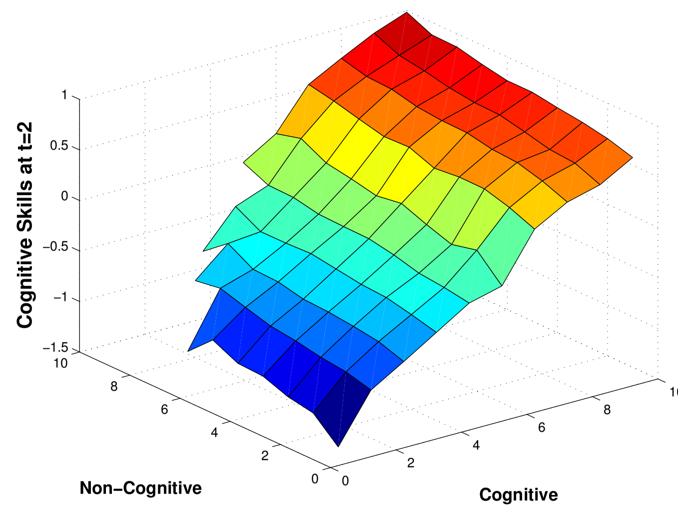
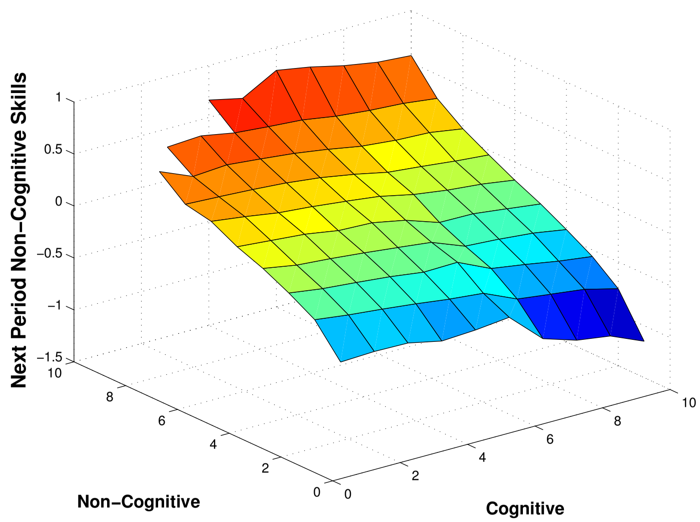
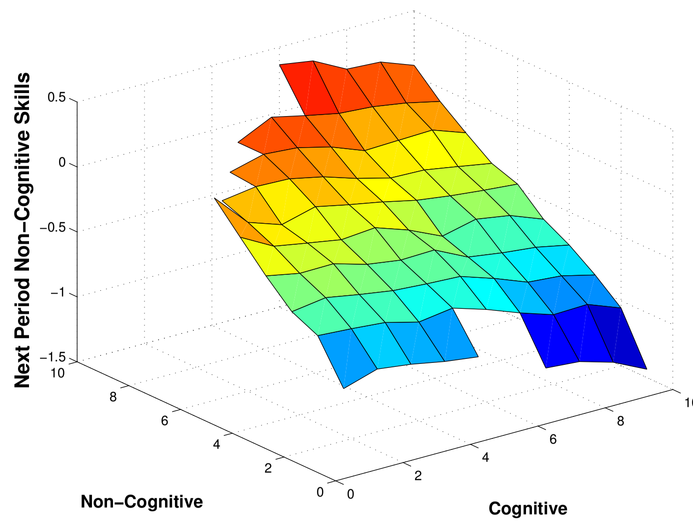


Fig. 2.8: θ^S_{t+1} as a function of θ^{NC}_t and θ^C_t $(t = 2)$

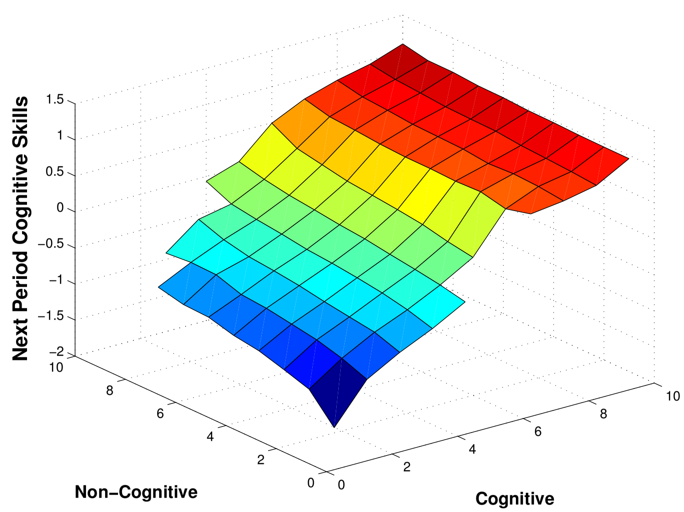
(a) Non-Cognitive: Everyone



(b) Non-Cognitive: Bullying Victims



(c) Cognitive: Everyone



(d) Cognitive: Bullying Victims

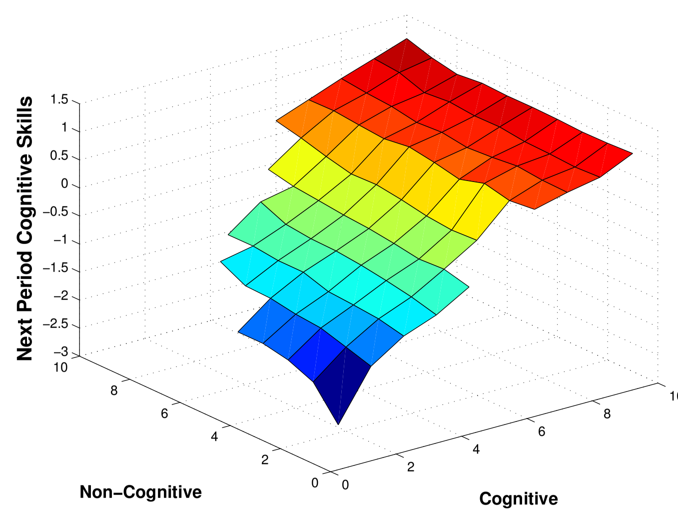
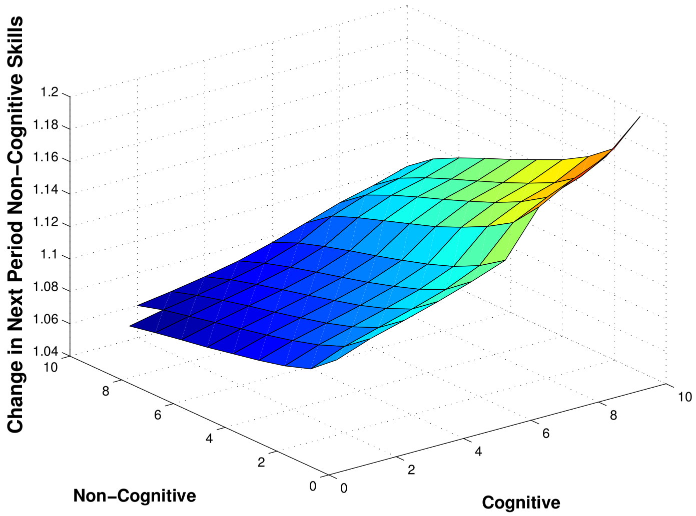
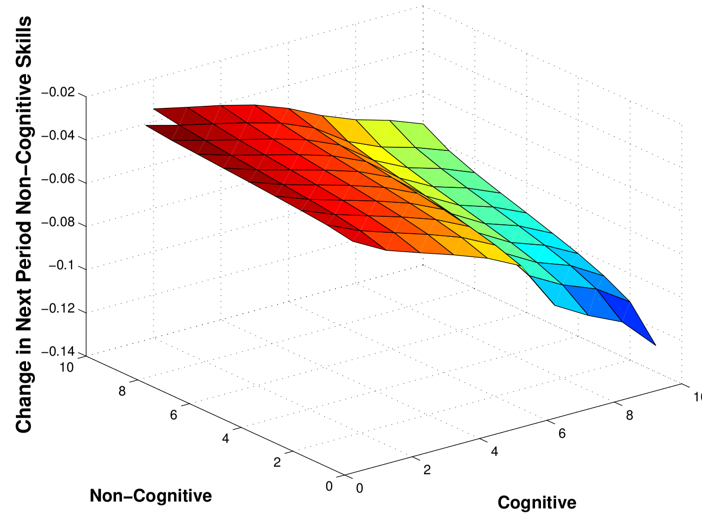


Fig. 2.9: Self $\left(\partial\theta^S_{t+1}/\partial\theta^S_t\right)$ and Cross $\left(\partial\theta^S_{t+1}/\partial\theta^{S'}_t\right)$ Productivities ($t=1$)

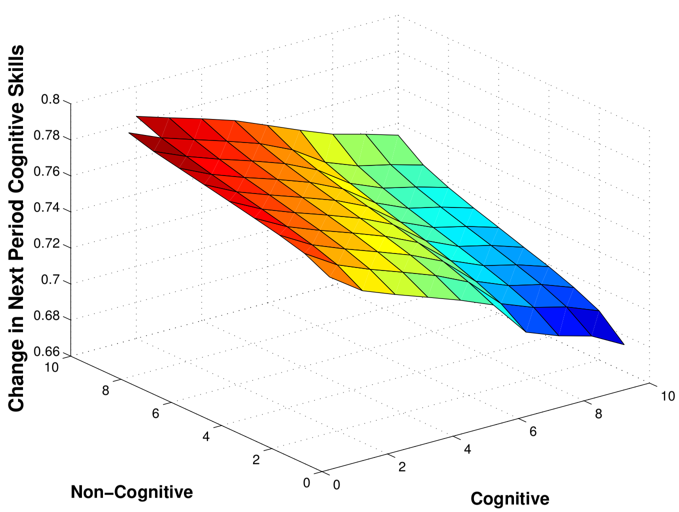
(a) Non-Cognitive: Self Productivity



(b) Non-Cognitive: Cross Productivity



(c) Cognitive: Self Productivity



(d) Cognitive: Cross Productivity

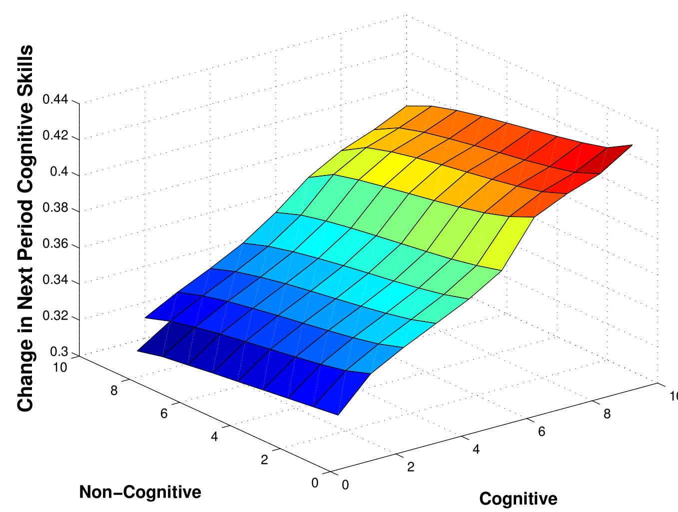
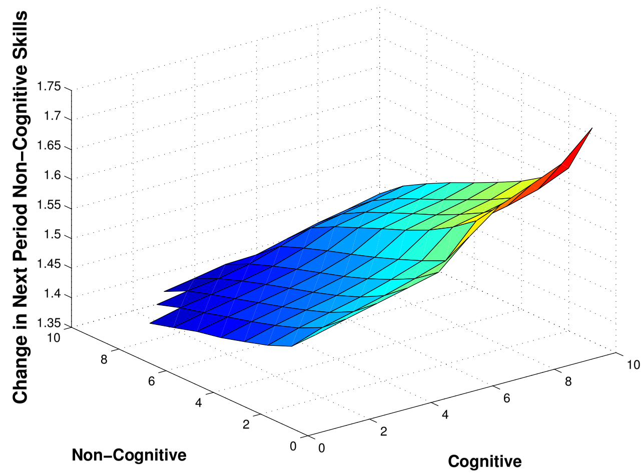
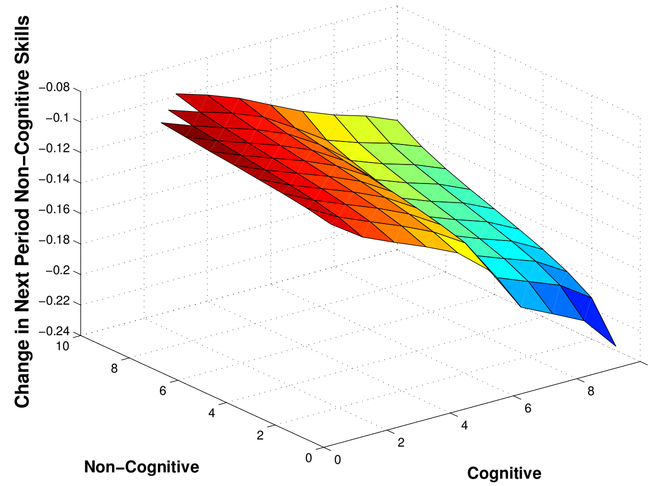


Fig. 2.10: Self $\left(\partial\theta^S_{t+1}/\partial\theta^S_t\right)$ and Cross $\left(\partial\theta^S_{t+1}/\partial\theta^{S'}_t\right)$ Productivities $(t = 2)$

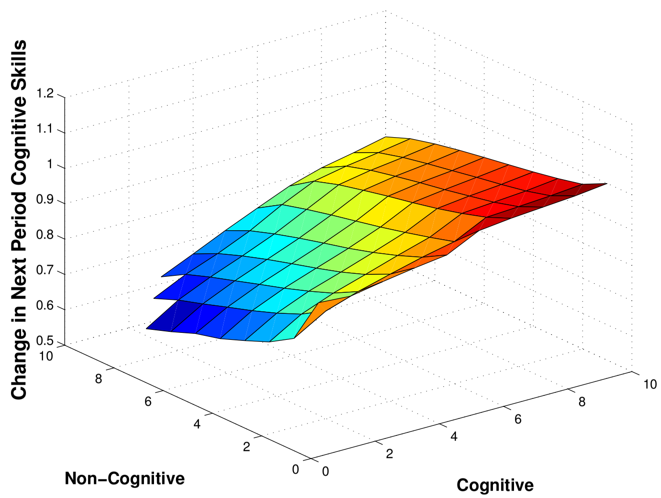
(a) Non-Cognitive: Self Productivity



(b) Non-Cognitive: Cross Productivity



(c) Cognitive: Self Productivity



(d) Cognitive: Cross Productivity

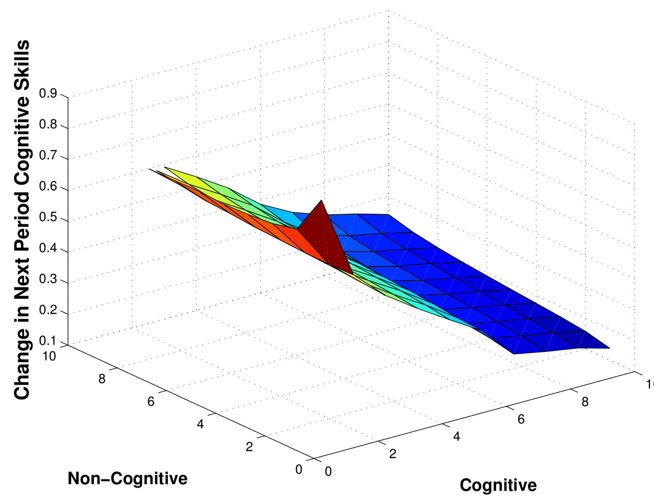
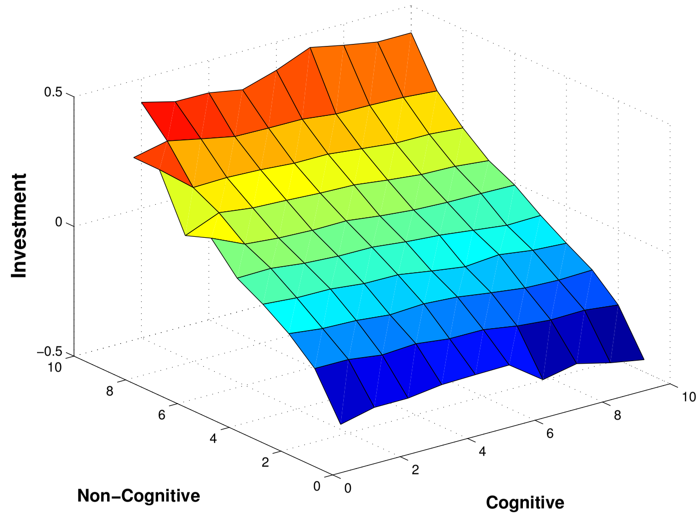
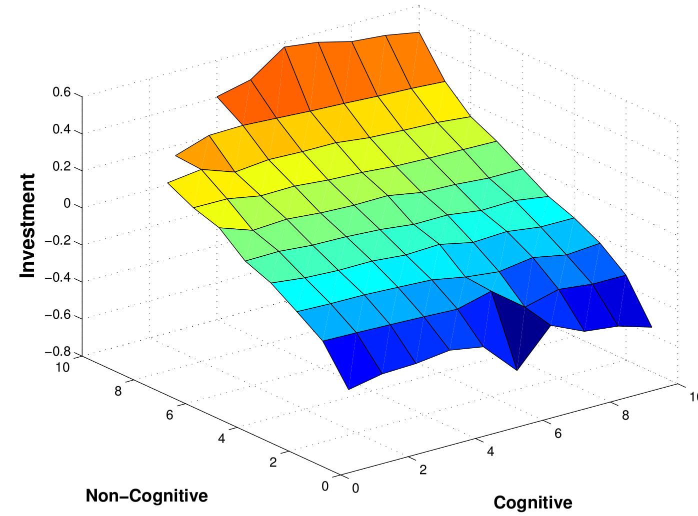


Fig. 2.11: Skill Investment as a function of θ_t^{NC} and θ_t^{C}

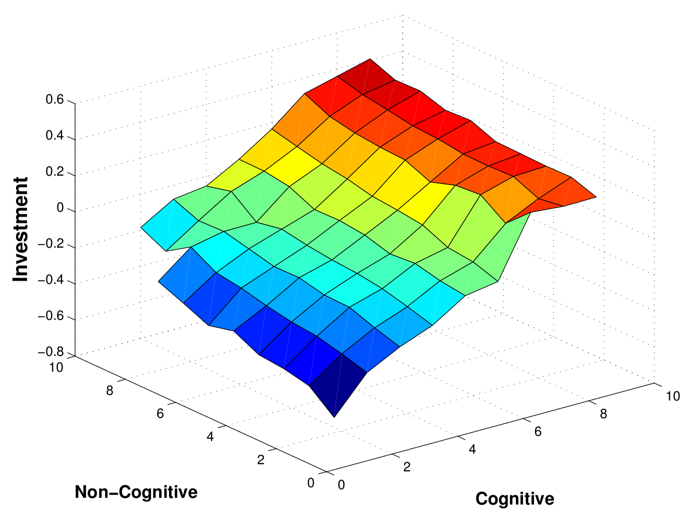
(*a*) Investment in Non-Cognitive Skills at $t = 1$



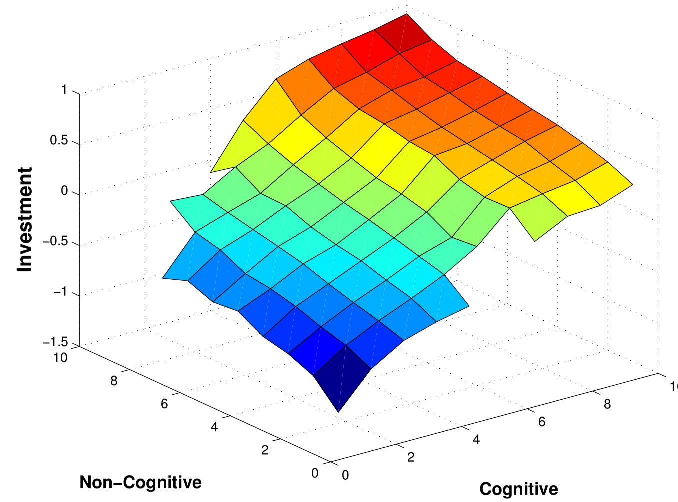
(*b*) Investment in Non-Cognitive Skills at $t = 2$



(*c*) Investment in Cognitive Skills at $t = 1$



(*d*) Investment in Cognitive Skills at $t = 2$



is presented in Figure 2.12a. It shows that this effect can be twice as big for youths with low initial levels of non-cognitive skills.

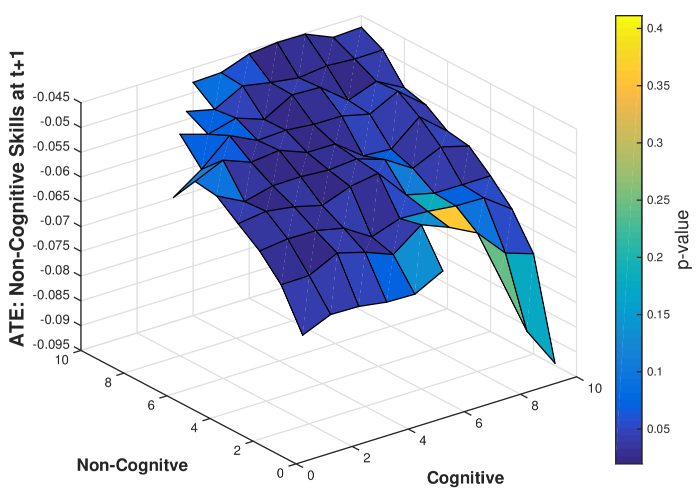
The same estimation shows there is no statistically significant effect of bullying on cognitive skill accumulation at that age. These results indicate that, as expected, bullying is much more costly in the non-cognitive dimension than in the cognitive one. Although victims might skip school, their learning ability is not affected as gravely as their ability to self regulate, overcome obstacles, see themselves positively or relate with others.

Due to the fact that skill levels are important determinants of future bullying, the documented skill depletion caused by bullying is translated into higher probabilities of being victimized again. My dynamic model allows me to see the difference in the incidence of victimization conditional on past victimization. That is, I am able to compare the probability of being bullied at $t = 2$ conditional on having being bullied at $t = 1$. The results presented in Figures 2.13a and 2.13b indicate that, despite the overall decrease of bullying incidence, previous victims are more likely to be bullied at $t = 2$ than previous non-victims regardless of their skill levels. The probability of being bullied at $t = 2$ of the previous non-victims is around one fourth less than that of the one that were previously victimized.

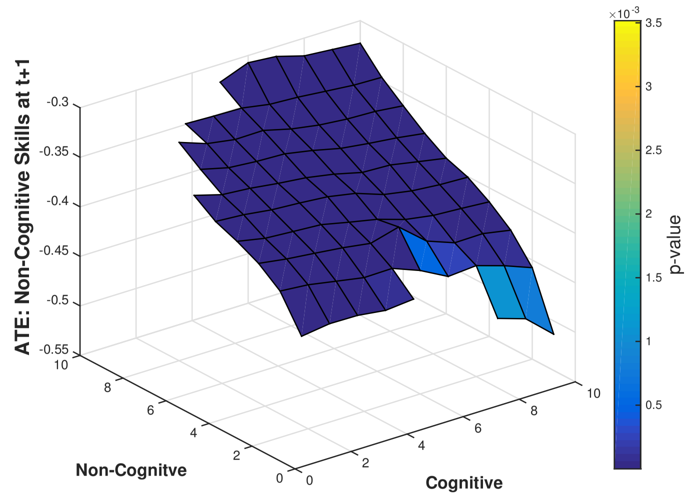
The dynamic analysis for age 15 (i.e., $t = 2$) finds that bullying not only becomes more selective, but also more costly. Its effect on the accumulation of non-cognitive skills now reaches -0.3763 for the average student, which represents a reduction of 72.8% of a standard deviation of next period skills. This implies loosing around half of a standard deviation in the language and math test scores. Figure 2.12b shows that at $t = 2$, bullying is also much more costly to those who lack non-cognitive skills.

Fig. 2.12: $E\left[\theta_{t+1}^{S}|\theta_t^{NC}, \theta_t^{N}, M_{t+1} = 1\right] - E\left[\theta_{t+1}^{S}|\theta_t^{NC}, \theta_t^{N}, M_{t+1} = 0\right]$

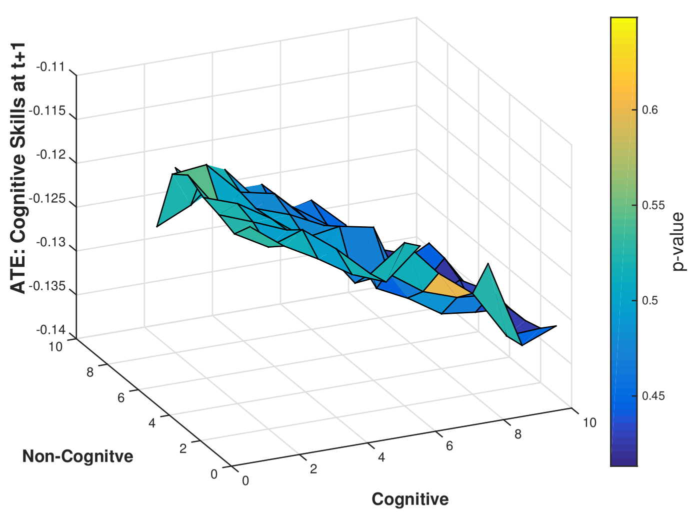
(a) Non-Cognitive $t = 1$



(b) Non-Cognitive $t = 2$



(c) Cognitive $t = 1$



(d) Cognitive $t = 2$

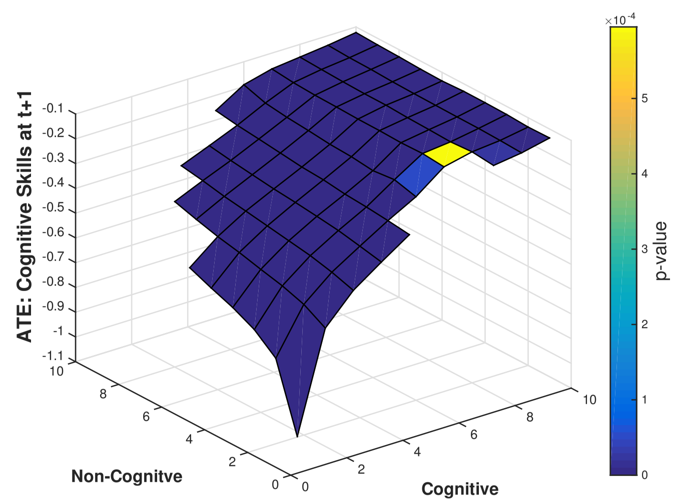
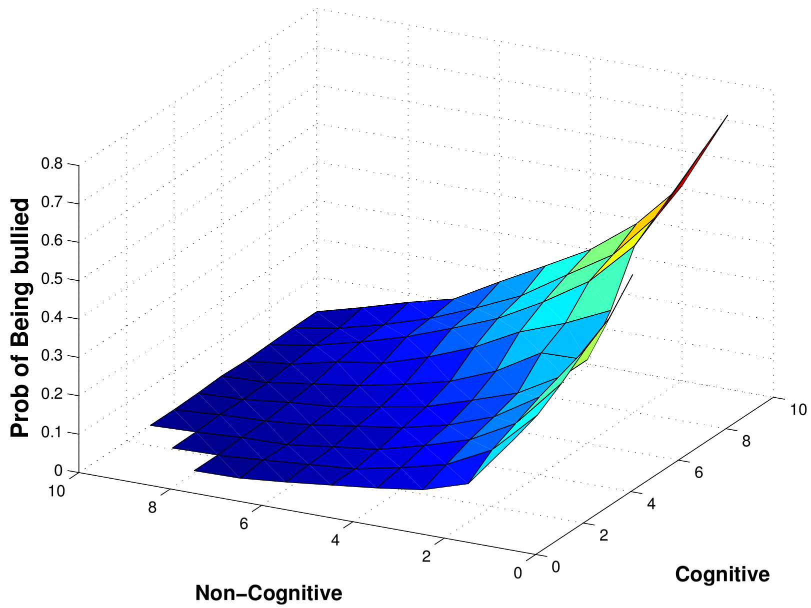
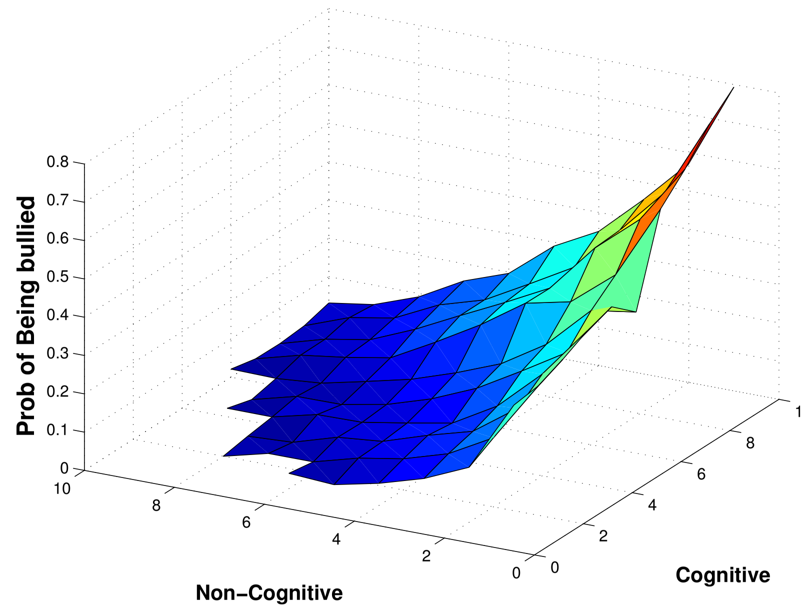


Fig. 2.13: Probability of being bullied at $t = 2$ depending on bullied or not at $t = 1$

(a) Not bullied at $t = 1$



(b) Bullied at $t = 1$



In particular, the stock of non-cognitive skills lost to bullying by the kids in the first decile of the non-cognitive skill distribution is around 1.5 times greater than the loss of those in the top decile of the distribution.

By age 15, and probably because bullying has became more selective, I do observe statistically significant effect on cognitive skill accumulation. In fact, I estimate that the average kid loses 35% of a standard deviation of next period cognitive skills. This represents loosing 16.64% of a standard deviation in the language test, and 13.42% of a standard deviation in the math test. As in the case of non-cognitive skill accumulation, the size of the effect depends on the skill levels at the beginning of the period. Figure 2.12d shows that the kids that arrive to this age with low cognitive skills feel the greatest impact of bullying on cognitive skill accumulation. In particular, the cognitive unskilled might lose up to 70% of a standard deviation in next period cognitive skills.

In sum, the average kid will score around two thirds of a standard deviation less in the language and math tests by age 16 if bullied during the last year. These skill loses can also be translated into sizable effect in terms of other outcomes besides test scores.[26] For instance, the average kid would be 28.6 percentage points more likely to report being sick recently, 8.3 percentage points more likely to smoke and 15.8 more likely to drink alcoholic beverages. The stock of skills lost also translates to setbacks in mental health. They equate to an increase of 134.5% of a standard deviation in the depression symptom scale, an increase of a full standard deviation in the levels of stress cased by insecurities regarding his or her image, and 95% of a standard deviation in the levels of stress caused by issues regarding school.

[26] In Appendix M, I run models of unobserved heterogeneity at age 16 on several outcomes like depression, stress in different situations, and the likelihood of smoking, drinking alcohol, felling healthy, being satisfied with life and going to college by age 19.

The full path for the average student. My model allows me to track all the process triggered by bullying and in that way measure the gap that widens as time goes by and new victimization events materialize. For instance, let us track the average individual, call it V. Being the average individual, V has an initial set of skills of $\left(\theta_t^{NC}, \theta_t^{C}\right) = (0,0)$. V has a 22.5% chance of being bullied. If V is bullied, V will lose 14.15% of a standard deviation of next period non-cognitive skills. V's cognitive skills remain unchanged. The fact that V was bullied increases V's probability of being bullied again by one fourth compared to the scenario when he was not bullied in the first place. If V is bullied again, let us call this scenario V_{11}, V would have lost 91.65% and 38.47% of a standard deviation of non-cognitive and cognitive skills respectively. If V is not bullied again but was bullied before, scenario V_{10}, V's skill loses would be of 17.23% of a standard deviation of non-cognitive skills and 2.8% of a standard deviation of cognitive skills. If V was not bullied before but is bullied now, scenario V_{01}, V's losses would amount to 72.8% of a standard deviation in the non-cognitive dimension and 35% of a standard deviation in the cognitive one. These results compare to the situation V_{00} when V was never victimized, and therefore V's set of skills remained at $(0,0)$.

Using the results presented in Appendix M to translate these gaps into more understandable metrics, we see that for V_{11} they represent a 36 percentage points increase in the probability of feeling sick, a 10.3 percentage points increase in the likelihood of smoking, 19.8 percentage points increase in the likelihood of drinking alcohol. They also represent an increase of 170% of a standard deviation in the depression symptom scale, an increase of 133% of a standard deviation in the stress due to image scale, and a 120% of a standard deviation increase in the stress caused by school.

Note that this exercise follows the average student. Therefore, around half of students in the sample will face harsher consequences than the ones just described due to the fact that they will start this middle-school journey with lower stocks of skills.

2.6.3.4 Complementarities

As explained in Section 2.3, the analysis of the static and dynamic complementarities between skills and bullying allows the measurement of how much the effect of the shock on skill formation is modified by a marginal change in previous period skills, and to what extent the effect of bullying is compounded on past bullying events. My results regarding the static complementarity on bullying (i.e., equation (2.2)), presented in Figures 2.14 and 2.15, show that marginal increases of the level of non-cognitive skills drastically reduce the negative effect of bullying on non-cognitive skill accumulation. In fact, comparing Figures 2.12a (i.e., the size of the effect on bullying at age 14 on non-cognitive skill formation at age 15) and 2.14a we see that a marginal increase in non-cognitive skills would have reduced the negative effect of bullying by around one half. For the average kid, the raw effect of bullying on non-cognitive skills at age 15 would be reduced from -0.0631, to -0.0368 by just a marginal increase in the previous period non-cognitive skills. At age 16, the static complementarity between bullying and non-cognitive skills represents around one third of the full effect of bullying on non-cognitive skill formation. Therefore, marginal increases in the previous period non-cognitive skills would have brought down the size of the bullying effect for the average kid from -0.3763 to -0.2414. These sizable reductions contrast with the evidence presented in Figures 2.14b and 2.15b that demonstrate the negligible influence that marginal increases of cognitive skills have on reducing the impact of bullying on non-cognitive skill accumulation.

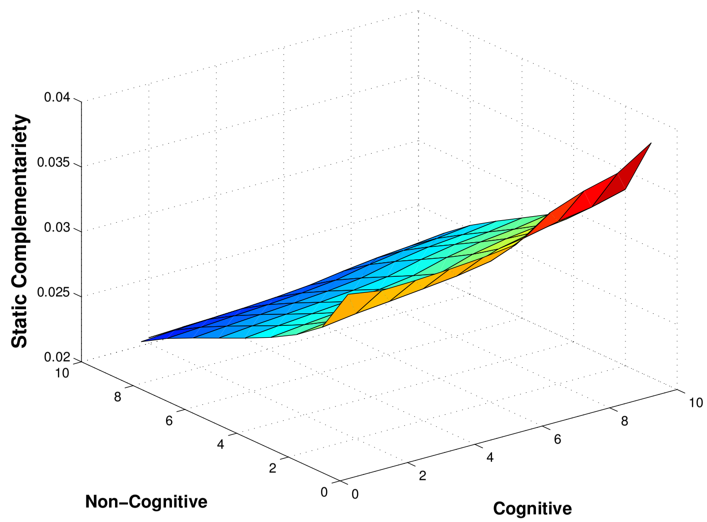
Regarding the effect bullying has at age 15 on cognitive skill accumulation, Figures 2.15c and 2.15d indicate that marginal increases in the stock of both cognitive and non-cognitive skills lessen its negative effect.[27] For the average kid, the effect of bullying on cognitive skill production would fall by a third due to marginal increases in previous period non-cognitive skills, and two thirds due to a marginal increase in the previous period stock of cognitive skills.

The analysis on the dynamic complementarity of bullying on skill accumulation (i.e., equation (2.3)), presented in Figure 2.16, shows that there is, in fact, a compounded effect of bullying, especially for those with low non-cognitive skills. That is, the negative effect of bullying at age 15 on skill accumulation at age 16 is greater if the person was bullied at age 14. Hence, not only bullying becomes more selective on low non-cognitive skilled people, as shown in Figures 2.6a and 2.6b, but also a second bullying event itself is more harmful.

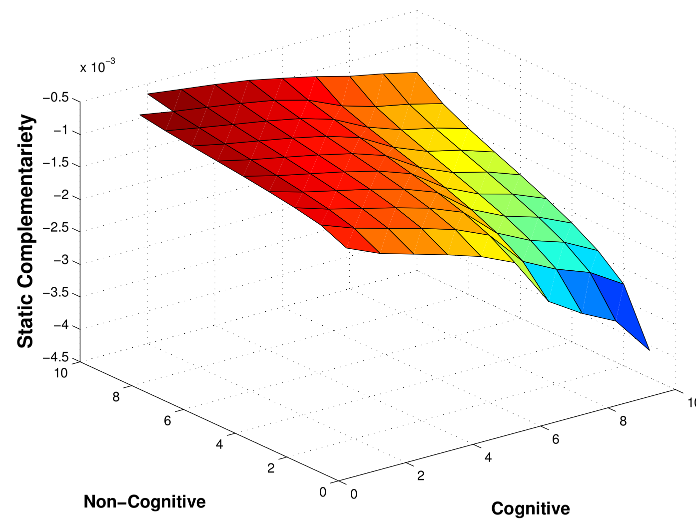
[27] Figures 2.14c and 2.14d present the static complementarity analysis for cognitive skill accumulation at age 14. They are presented for the sake of completeness. I will not thoroughly analyze their results because there was no effect of bullying on that skill dimension for that age in the first place.

Fig. 2.14: Static Complementarity at $t = 1$

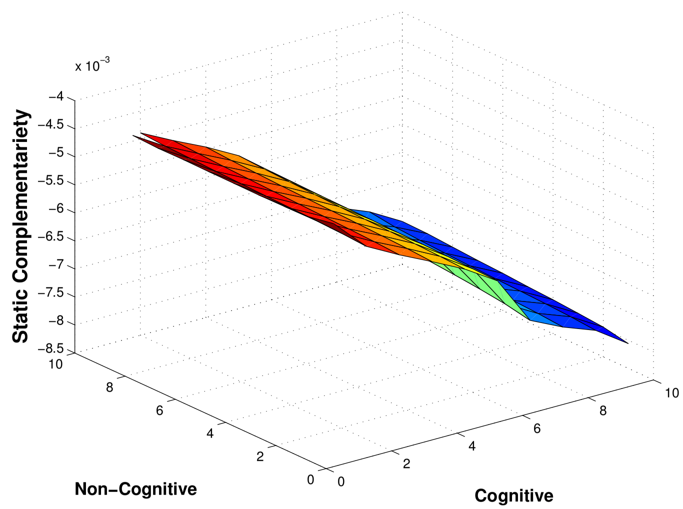
(a) $\partial\left(\frac{\Delta\theta_{NC,t+1}}{\Delta M_{t+1}}\right)/\partial\theta_{NC,t}$



(b) $\partial\left(\frac{\Delta\theta_{NC,t+1}}{\Delta M_{t+1}}\right)/\partial\theta_{C,t}$



(c) $\partial\left(\frac{\Delta\theta_{C,t+1}}{\Delta M_{t+1}}\right)/\partial\theta_{NC,t}$



(d) $\partial\left(\frac{\Delta\theta_{C,t+1}}{\Delta M_{t+1}}\right)/\partial\theta_{C,t}$

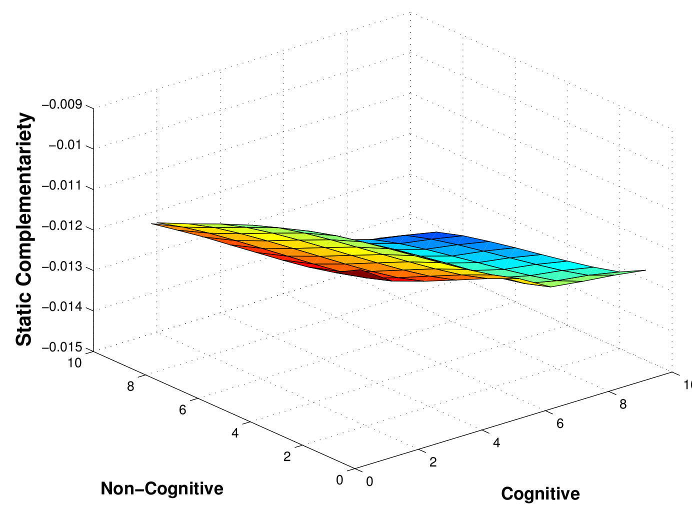
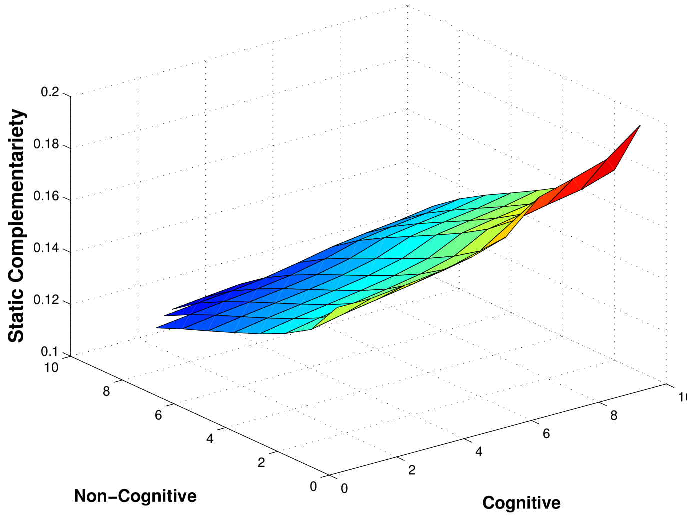
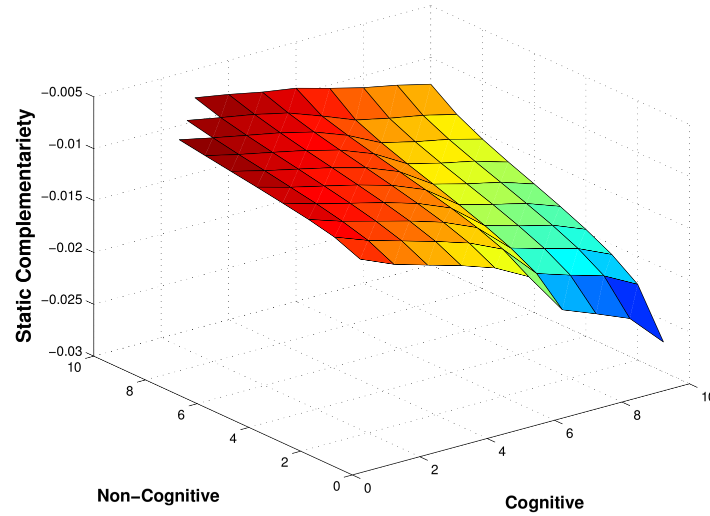


Fig. 2.15: Static Complementarity at $t = 2$

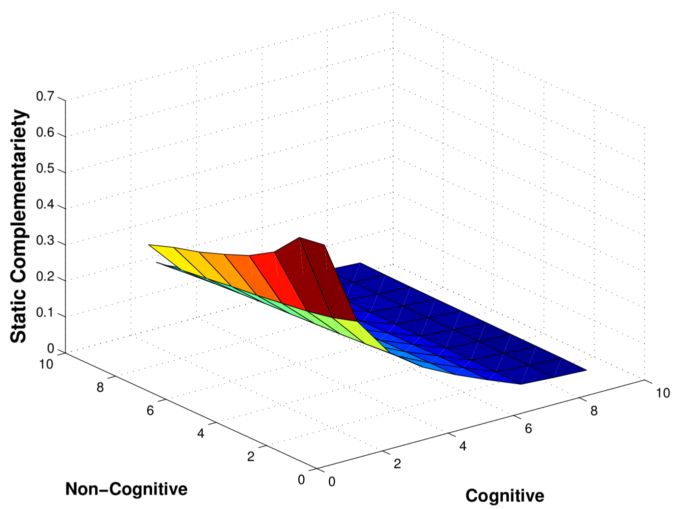
(a) $\partial\left(\frac{\Delta\theta_{NC,t+1}}{\Delta M_{t+1}}\right)/\partial\theta_{NC,t}$



(b) $\partial\left(\frac{\Delta\theta_{NC,t+1}}{\Delta M_{t+1}}\right)/\partial\theta_{C,t}$



(c) $\partial\left(\frac{\Delta\theta_{C,t+1}}{\Delta M_{t+1}}\right)/\partial\theta_{NC,t}$



(d) $\partial\left(\frac{\Delta\theta_{C,t+1}}{\Delta M_{t+1}}\right)/\partial\theta_{C,t}$

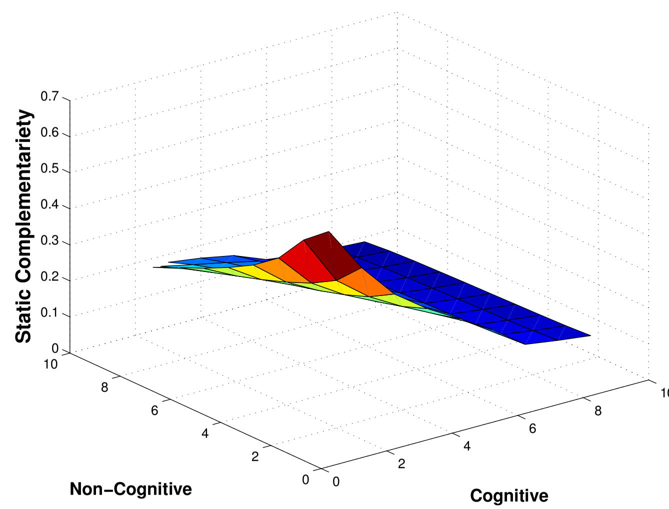
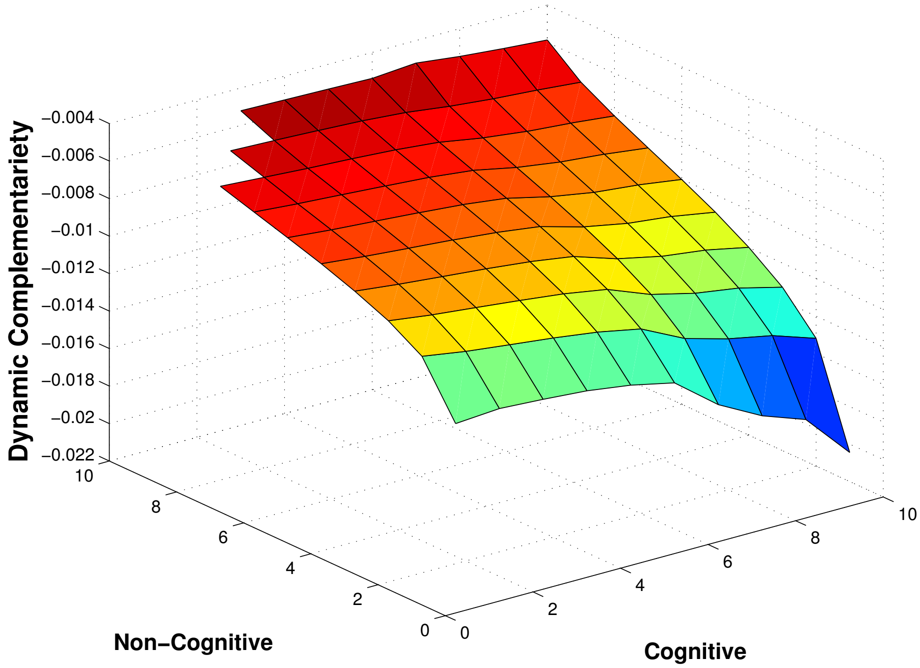


Fig. 2.16: Dynamic Complementarity



All the evidence presented in this Chapter argue in favor of the existence of a self-reinforcing mechanism in which low skilled kids are more likely to be victims of violence at their schools, and in turn not only their skills are depleted by the bullying event itself, but also its consequences aggravated for those who started with low skill levels in the first place. This send them in a downward spiral by making them even more at risk of being victims of bullying in the future, which in turn will be much more harmful events, and therefore having always more and more difficulties in acquiring the non-cognitive skills they lack. Even though, I show that investment in non-cognitive during middle school years is often unproductive, the static complementarity results suggest that even a tiny bit of skill accumulation would have an immense impact not only in deterring bullying, but also in lessening its consequences among those that are more at risk.

2.7 Policy Implications

Several anti-bullying campaigns have been deployed all around the world in an effort ambitious effort to eliminate this unwanted phenomenon. Prominent psychologists and governmental institutions are continuously involved in the development of programs to deter bullying.[28] My findings indicate there are at least two fronts policy-makers can work on. First, the development of non-cognitive skills. Non-cognitive skilled kids will not only be more likely to be successful adults (Duckworth and Seligman, 2005; Heckman et al., 2006c), but also they are dramatically less likely to be victimized. And if—despite the low probability of being so— they happen to be bullied, its impact on their skill accumulation path is much lessened. The importance of developing non-cognitive skills at a young age is heightened by the strong dependance of current skill levels on past skills levels.

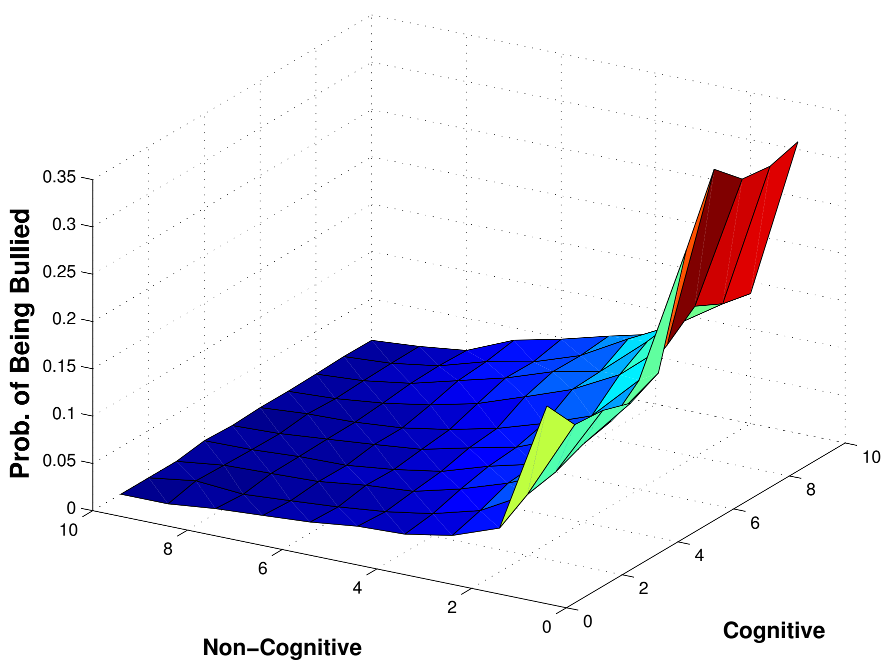
The second implication of my results relates with classroom assignment. Tables 2.6 and 2.7 show that, given the skill levels, children with uncommon characteristics are more likely to be targeted by bullies. Therefore, in an effort to illustrate how much would bullying be reduced if kids were less likely to be found in classroom where some of their characteristics end up being uncommon, I simulate a different type of classroom allocation. One that is unfeasible in practice and treated as a benchmark, places all the kids in the survey in classroom with kids that have similar stocks of non-cognitive skills. This exercise ignores geographical distances. It just sorts the universe of students with respect to their non-cognitive skills and split them in classrooms according to the typical classroom size in South Korea.

The results of these simulations are presented in Figure 2.17. As in Figure 2.6a,

[28] See the Olweus Bullying Prevention Program and the US Education Department stopbullying.gov program.

it plots the likelihood of being bullied for every skill level. A comparison between these two figures shows the massive impact that reducing in-classroom non-cognitive skill heterogeneity has on the likelihood of being victimized. The benchmark case presented in Figure 2.17 shows that by arranging students with classmates that have similar levels of non-cognitive skills, the overall likelihood of victimization falls from 22.5% to 5.5%. This dramatic reduction is focused on medium and high non-cognitive skilled students, for which the hazard of being bullied almost disappears. Although proportionally less, there is also a substantial reduction of victimization among low non-cognitive students. If placed in skill homogenous classrooms, low non-cognitive skilled students would see their chances of being victimized drop by half.

Fig. 2.17: Classroom Allocation Simulations: Benchmark



2.8 Conclusions

This chapter of the dissertation develops and estimates a structural model of skill accumulation that explicitly introduces a peer-affected input. This way, I introduce bullying as a skill depleter event into a model that is general enough to allow past skills, parental investment choices and bullying to affect future stock of skills, and at the same time investment choices and bullying to be affected by past skill levels. The model uses several dimensions of unobserved heterogeneity and in-classroom variation of student characteristics to identify the victims. My findings indicate the existence of a vicious cycle between victimization and skill depletion. I find that bullying is disproportionately suffered by students that lack socio-emotional skills, and among those, the smart students are more likely to be victimized. My findings, in line with psychological studies, suggest that conditional on the level of skills, kids with uncommon characteristics relative to those of their classmates are more likely to be victimized.

The dynamic estimation showed that bullying is very costly in terms of the amount of skills lost from one period to the next. Bulling at age 14 reduces non-cognitive skill accumulation by a 14.15% of a standard deviation for the average kid. That effect is twice as big for kids with low initial levels of non-cognitive skills. At age 15, bullying not only becomes more selective, but also more costly. It reduces next period skills by almost three-fourths of a standard deviation. Static complementarity shows current stock of non-cognitive skills—unlike the cognitive one—influences greatly the "negative productivity" of the bullying event. In the same vein, the analysis of dynamic complementarity shows that the effect of being bullied at age 15 on the skills at age 16 is greater for those who were also bullied at age 14, especially those who started with low non-cognitive skills.

These results show the existence of a self-reinforcing mechanism, in which initial levels of skill becomes crucial, suggesting that policies aimed to foster cognitive skills at early ages will greatly reduce victimization occurrence. In addition, my model indicates that the allocation of students in more homogeneous classroom might reduce victimization by preventing kids with uncommon characteristics to be isolated and targeted by bullies.

This Chapter intends to contribute to the human development literature in economics with explanations of how victimization of school-aged kids may hamper the development of successful adults. In this process, this Chapter contributed to the skill formation literature by introducing latent factor dependent shocks as triggers of phenomena that have long-lasting consequences. In this context, this Chapter opens a promising research agenda that can continue in at least two ways. First, by extending my model to incorporate additional characteristics and dynamics regarding bullying, and second, by analyzing the skill accumulation consequences of other types of shocks. Among the latter, new research is needed to analyze the consequences of shocks like parental separation or loss, and health or family financial mishaps. Among the former, promising research opportunities arise in the analysis of the role that gender plays in classroom dynamics *vis-a-vis* bullying, or—data permitting—the introduction of physical traits as determinants of victimization. Furthermore, given the importance of initial levels of skills, it is crucial to explore how bullying affects skill accumulation among younger cohorts.

3. WAGE DISCRIMINATION AGAINST HOMOSEXUALS, *THE ROLE OF SKILLS*

3.1 Introduction

Labor market discrimination has been a popular topic in economics. In particular, labor market discrimination against women and racial minorities has been under the radar of labor economists for a long time (see for example, Becker, 1957; Mincer and Polachek, 1974; Neal and Johnson, 1996; Urzua, 2008). Their findings have ignited a number of anti-discriminatory policies around the world.[1] However, much less is known regarding labor market discrimination towards sexual minorities, despite the increasing amount of public attention received by this population in recent years.[2] Recent literature has tried to keep up with the public debate and has been able to document the existence of a sexual minority wage gap. In particular, this literature has shown that non-heterosexual men earn less than their heterosexual counterparts by about 15–30 percent (Badgett, 1995; Black et al., 2003; Carpenter, 2007). Evidence on women is less conclusive. While some studies like Badgett (1995) suggest gay/bisexual women earn less than heterosexual ones, other studies like Berg and Lien (2002), Black et al. (2003) and Antecol et al. (2007) suggest the opposite.

[1] In the US only, laws prohibiting labor market discrimination based on race, religion, country of birth, gender, age and mental disability have been passed (Clain and Leppel, 2001).

[2] Recent years have seen an unprecedented push of the political agenda regarding the rights of sexual minorities in terms of employment protection, marriage and adoption(Carpenter, 2009).

There are at least two major problems with the existing literature on sexual minorities' labor market discrimination. First, they do not deal with the fact that schooling and occupational choices are endogenous and themselves possibly affected by the sexual preferences as well.[3] And second, the income gap might respond to unobservables, namely skill differentials, not only in productivity terms but also in the way people choose their schooling level or occupation. This Chapter overcomes these two issues by using an empirical strategy that controls for unobserved skill differentials, and based on them allows choices and labor market outcomes to be endogenous. This is a significant improvement with respect to the existing literature, as schooling and occupational decisions are often included in the analysis as controls either by assuming they are exogenous variables (Badgett, 1995), or by completely ignoring the endogeneity problem (Allegretto and Arthur, 2001; Black et al., 2003).[4]

Another problem that plagues this literature is the lack of appropriate data sets. Papers on sexual-preference discrimination have relied mainly on two sources of data: the General Social Survey (GSS) or the Census.[5] Each has its own problems. On one hand, the GSS provides a relatively detailed description of sexual preferences but has a very reduced sample size. For instance, the sample used by Badgett (1995) comprises only 34 non-heterosexual women and 47 non-heterosexual men. On the other hand, the Census provides a greater sample size, but the definition of who belongs to a sexual minority has important setbacks. The only way to do so is to identify people

[3] In fact, Carpenter (2009) shows that college outcomes are different for non-heterosexual men, and Allegretto and Arthur (2001) find that—cohabiting—homosexual males have significantly higher levels of education than heterosexuals. In addition, Antecol et al. (2007) shows that this finding is also true for—cohabiting—homosexual women.

[4] Black et al. (2003) in fact discuss the endogeneity problem of occupational choice but ignores the one related to education.

[5] Carpenter (2007) uses the National Health and Nutrition Examination Surveys (NHANES III), but faces the same problem as the papers that use the GSS.

who cohabit with a person of the same gender (Clain and Leppel, 2001; Allegretto and Arthur, 2001; Antecol et al., 2007). Therefore, those identified as non-heterosexuals conform a selected sample that excludes people who have chosen to remain single. The cultural and legal restrictions imposed to gay marriage give the grounds to claim that the selection into marrying/cohabiting is very different for homosexuals than for heterosexuals, and that it may rely heavily on unobservables that also affect the schooling and occupational choices, and the wages earned. I overcome these problems by using a novel longitudinal Norwegian data set that provides very detailed information regarding sexual preferences, allows me to have an acceptable sample size, and importantly, allows me to measure skills as a latent variable.[6] This way, I do not have to restrict the study to non-single homosexuals and I am able to avoid the source of confoundedness that arise from such a selected sample.

Another important contribution of this Chapter is that it is able to inquire about the extent to which the wage gaps observed respond to skill differences. Recent literature has shown that skills are critical in determining educational and labor market outcomes (Heckman et al., 2006a; Borghans et al., 2008; Espinoza et al., 2014), but whether there is a skill gap between homosexuals and heterosexuals, and whether that gap may explain the wage differential is unknown. A priori, there are no reasons to believe such a skill gap may exist, especially in the cognitive dimension. However, Chapter 2 in this dissertation finds that teasing or bullying have a significant impact in reducing skills, especially the non-cognitive ones. Then, if members of sexual minorities are more likely to be teased or bullied while young, they may face adulthood with a diminished stock of skills.

[6] For a detailed explanation of the data used to define the non-heterosexual groups see Section 3.2.1, and a more detailed explanation of the data in general see Section 3.3.

Finally, this Chapter also addresses whether the income gap responds to differences in occupational tastes. That is, if non-heterosexual people are more likely to choose lower-paying occupations, that choice may contribute in explaining the income differences between gays and heterosexuals. Unlike Antecol et al. (2007) who explore this issue—and find no evidence in its favor—under the assumption of occupational choices being exogenous, I allow them to be endogenous in the context of a Roy Model of potential outcomes based on observable and unobservable characteristics.

3.2 Background

3.2.1 Discretizing the Definition of Homosexuality According to the Data

Section 3.3 shows that the Norwegian dataset used in this Chapter has a very rich battery of questions regarding sexual orientation. However, regardless of the data available, defining homosexuality is a difficult task. Homosexuality is not a dichotomous indicator but a spectrum of preferences that go from being attracted or having had sex with only women to being attracted or having had sex with only men, where between of these two extremes there is a complex assortment of sexual preferences (Laumann et al., 1994). Defining homosexuality according to the data becomes even more complex as it can be drawn from three different concepts explored in Laumann et al. (1994) and Badgett (1995). First, which gender generates sexual interest (i.e., sexually attracted to, sexual fantasies about) to the respondent. Second, whether the respondent considers herself homosexual to a particular extent. And third, the gender of the respondent's sexual partners. In consequence, I build four types of dichotomous variables that split the sample between heterosexual and non-heterosexual respondents. The first measure is constructed according to sexual interests (SI) and considers to be homosexual any individual that is attracted in any degree to people of the same sex. The second measure is constructed according to self-perception of homosexuality (SP) and considers as homosexual any individual that believes he or she is homosexual in any degree. The third measure (SX) respond to whether a person has had sex with someone from the same gender. And finally, the fourth measure widens the definition of the third measure to also include those that are mainly attracted to people from the same gender, regardless of the gender of their sexual partners (SX+).[7]

Tab. 3.1: Percentage of Homosexuals According to Each Definition

	SI	SP	SX	SX+
Observations	644	591	350	528
% of Population				
Overall	22.0	20.2	11.9	18.0
Males	11.4	12.9	10.2	12.5
Females	30.7	26.1	13.3	22.4

Table 3.1 shows the proportion of respondents that are classified as homosexual according to each definition. It shows that the SI and SP definitions are, as expected, very similar to each other, while the definition that relies on actual sexual relations is much more selective. Interestingly, for every definition women are more likely to be classified as homosexuals, but the difference in the number of men and women identifies as homosexual is drastically smaller in the SX definition than in the ones that use self-perception or sexual interest.

[7] This last definition intends to increase the sample size while still having a very restrictive target of the population identified.

3.2.2 The Income Gap

The typical estimation found in the literature on income discrimination for sexual minorities relies on a linear model for labor market outcomes of the form

$$\ln Y = \alpha + \gamma Gay + \beta \mathbf{X} + \sum_{s=1}^{S} \phi_s D_s + \varepsilon \tag{3.1}$$

where Y is the income measure of interest, $\mathbf{X}$ is a set of observable controls (e.g., age, experience, geographical region), and D_s are dummy variables that represents the schooling level where $s = \{0, \ldots, S\}$.[8] To capture differences between the effect for lesbians from the effect of homosexual men, Model (3.1) is often augmented with an interaction between Gay and gender, or the models are estimated separately for each gender. In essence, the parameter of interest of these type of models is γ which the existing literature interprets as the difference in income between two individuals that have the same observable characteristics and have achieved the same education levels but differ in their sexual preferences.

However, the estimation of (3.1) requires substantial econometric considerations (Urzua, 2008). The most important one is that unobservable characteristics may influence simultaneously schooling or occupational decisions and income, and those choices themselves can be influenced by the sexual preference of the individual.[9] This issue has completely being ignored by the literature on homosexual-heterosexual wage gaps. Neal and Johnson (1996)—when exploring racial wage gaps—argue they can bypass that endogeneity issue by referring to the skills accumulated after schooling

[8] Berg and Lien (2002) use a multinomial logit because the income on the GSS data is coded in ranges. The intuition and controls used remain the same as in (3.1)

[9] Antecol et al. (2007) show that homosexuals are more likely to go to college and also that gay males are less likely to be in male dominated occupations than heterosexual men, while lesbians are more likely to do so than heterosexual women.

is no longer mandatory instead of schooling or occupation *per-se*. For them, the educational and occupational choices made during adulthood that end up affecting income are mediated by the stock of skills—proxied by test scores—they had when they started making those decisions. Hence, this approach claims that conditional on the skill levels, the income gap is fixed across educational levels or occupational sectors. As Urzua (2008) showed in the case of racial income gaps, I show that this approach is not appropriate to describe the existing income gaps between homosexuals and heterosexuals. In fact, the size of these gaps differ across educational and occupational choices.

Models like (3.1) also assume that ϕ_s does not differ across sexual preferences. Although, we are not interested in the estimates of ϕ_s, this assumption does affects the estimate of γ because of the endogeneity issues just described.

My empirical strategy takes care of these issues. It not only incorporates unobserved skill into the analysis of sexual preference related wage gaps, but also—based on these unobservables—allows schooling and occupational choices and labor market outcomes to be endogenous. It also allows homosexuals and heterosexuals to have a different set of mincerian parameters, eliminating the troublesome assumption of equal returns to observable characteristics and choices.

For comparison purposes, Table 3.2 presents the reduced form results of models of the type described by equation (3.1) for each definition of homosexuality. Column 1 presents the unconditional difference in means. Norwegian homosexuals, on average, earn between 11% and 15% less that heterosexuals. Adding age and gender controls in column 2 reduce this gap to between 7.3% and 13.8% depending on the homosexuality definition. However, the specification that most resembles the one used in the existing literature is that of column 3 with the introduction of a dummy variable that takes the value of 1 of the respondent had any kind of tertiary education. Also, it incorporates an interaction term between gay and female to identify a differential effect on lesbians. The income gap in this specification ranges from 9% to 19% depending on the homosexuality definition. However, for the most selective definitions the gap is no longer statistically different from zero at the 90% confidence level.[10] Compared to the results found in the US labor market, the Norwegian results seem somewhat small. This could respond to differences in the type samples used in the analyses or to structural features in the Norwegian labor market *vis-a-vis* the US. Column 4 of Table 3.4 presents the estimation results of a model *a la* Neal and Johnson (1996). That is, introducing test scores $\mathbf{T}$ as proxies of abilities instead of the schooling attainment:

$$\ln Y = \alpha + \gamma Gay + \beta \mathbf{X} + \lambda \mathbf{T} + \varepsilon$$

The results are very similar to those obtained with the college dummy instead of test scores.

Interestingly, there seems to be no differential effect for lesbians as the interaction term between *Gay* and female is never statistically different from zero. This is a very relevant result that should be taken into account later on as the structural model I fit in this Chapter is unable to disentangle differential effects for homosexuals depending on their gender, due to sample size restrictions.

[10] In fact, the coefficient for SX definition is statistically significant at the 85% confidence level.

Tab. 3.2: Gap in Yearly Income

Definition	(1)	(2)	(3)	(4)
SI	-0.121***	-0.098**	-0.189**	-0.184**
	(0.046)	(0.046)	(0.087)	(0.091)
SI×*female*			0.131	0.126
			(0.103)	(0.107)
SP	-0.136***	-0.095**	-0.160*	-0.163*
	(0.048)	(0.048)	(0.083)	(0.088)
SP×*female*			0.094	0.115
			(0.107)	(0.145)
SX	-0.156***	-0.138**	-0.126	-0.136
	(0.058)	(0.057)	(0.088)	(0.093)
SX×*female*			-0.025	-0.008
			(0.120)	(0.120)
SX+	-0.112**	-0.073*	-0.092	-0.084
	(0.049)	(0.048)	(0.082)	(0.086)
SX+×*female*			0.016	0.045
			(0.106)	(0.106)
Age & gender		X	X	X
College			X	
Test Scores				X

Note: Standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1. All estimations were drawn from different regressions with the dependent variable being the natural logarithm of yearly income among the economically active population in wave four of the survey. Age enters in a quadratic polynomial. The College variable takes the value of 1 when the person has undertaken any kind of tertiary education in colleges (*høyskole*) or universities (*universitetet*). This includes three-year, four-year college studies or studies in the university with the same length. The estimations with test scores include six scores, one related to school grades (Math, Norwegian and English), two of them relate to agreeableness and the ease to make friends, the others relate to positivism, leadership and scholastic competence. All of them collected in the first wave of the survey. Estimations including college and test scores we also calculated with no different results form those of column (3) or (4).

3.2.3 Access to Tertiary Education

Norway has a relatively high tertiary education enrollment rate; around 62% of people go to a tertiary education institution of some sort. A vast literature on returns to schooling has shown that differences in access to tertiary education is critical in determining difference in earned income (Mincer, 1958; Becker and Chiswick, 1966; Card, 2001). Therefore, part of the sexual preference income gap could be explained by differences between homosexuals and heterosexuals in enrollment rates in tertiary education. However, the reduced-form estimations presented in Table 3.3 show that there is no systematic evidence of either group being more likely to go to a tertiary education institution.[11] If anything, these results suggest that lesbians tend to go less to college than heterosexual women. Interestingly, when statistically different from zero, the negative effect on lesbians counters the positive effect on women—not presented in the table. Hence, the probability lesbians of going to tertiary education is the same as that of homosexual and heterosexual men.

[11] These findings oppose results on US data that suggests that homosexuals are more likely to reach college education (Carpenter, 2009).

Tab. 3.3: Gap in College Attendance

Definition	(1)	(2)	(3)	(4)
SI	0.068***	0.032	0.041	0.024
	(0.025)	(0.025)	(0.048)	(0.048)
SI×*female*			-0.013	0.001
			(0.056)	(0.056)
SP	-0.008	-0.028	0.039	0.023
	(0.026)	(0.026)	(0.046)	(0.046)
SP×*female*			-0.099*	-0.051
			(0.055)	(0.055)
SX	-0.016	-0.016	0.067	0.072
	(0.032)	(0.031)	(0.049)	(0.049)
SX×*female*			-0.137**	-0.139**
			(0.063)	(0.063)
SX+	-0.083***	-0.094***	0.018	0.025
	(0.027)	(0.026)	(0.045)	(0.045)
SX+×*female*			-0.169***	-0.150***
			(0.056)	(0.055)
Age & gender		X	X	X
Test Scores				X

Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The College variable takes the value of 1 when the person has undertaken any kind of tertiary education in colleges (*høyskole*) or universities (*universitetet*). This includes three-year, four-year college studies or studies in the university with the same length. Age enters in a quadratic polynomial. The estimations with test scores include six scores, one related to school grades (Math, Norwegian and English), two of them relate to agreeableness and the ease to make friends, the others relate to positivism, leadership and scholastic competence. All of them collected in the first wave of the survey. Sample comprises the economically active by wave 4 to be comparable to structural model results.

3.2.4 Occupation

Occupational sorting is another important determinant of wages. There are occupations that inherently pay on average more than others. Then, inquiring if homosexuals tend to prefer low paying occupations more than their heterosexual counterparts is key in addressing income discrimination. In fact, there is evidence of occupational sorting among sexual minorities. Antecol et al. (2007) find evidence for the US that suggest that gay males are less likely to be in male dominated occupations, while lesbians are more likely to do so. However, according to Antecol et al. (2007), this particular kind of sorting has no role in explaining the income gap.

The reduced form results presented in Table 3.4 provide evidence in favor of the hypothesis that Norwegian homosexuals—especially those of the broader definitions—prefer blue collar over white collar occupations, and as blue collar occupations pay on average less, this could explain part of the income gaps observed.

3.3 Data

This Chapter of the dissertation uses a novel data set that has not been explored by the economic literature. The Young in Norway (YiN) is a longitudinal data set product of a research project headed by the Program for Adolescent Research (Ungforsk) in 1990.[12] YiN is an amazing effort to collect detailed information about the characteristics, attitudes and choices of a cohort of Norwegians. In particular, it has four features that makes it the best data set available to study sexual minority income discrimination. First it is longitudinal, therefore I am able to observe respondents' characteristics when young. This improves over existing literature that explores income gaps of sexual minorities using non-longitudinal surveys, which in turn prevents researchers from differentiating away time-invariant unobservables that might bias

[12] The project had four main goals. First, it was designed to produce studies of the entire population, not just specific groups. For that, the study's design needed to have a high response rate and national representativity. Second, the study wanted to provide researchers tools to find causal relations. Finally, the study had to be comprehensive enough to be able to provide information of adolescents across different social arenas (Strand and von Soest, 2008).

Tab. 3.4: Difference in White Collar Occupations

	(1)		(2)		(3)		(4)	
Def.	Coeff.	Std.Err.	Coeff.	Std.Err.	Coeff.	Std.Err.	Coeff.	Std.Err.
SI	-.0053	(0.024)	-.0099	(0.024)	-.0823**	(0.047)	-.1132***	(0.046)
SP	-.0938***	(0.025)	-.085***	(0.025)	-.1142***	(0.043)	-.1064***	(0.044)
SX	-.0234	(0.031)	-.0205	(0.03)	.005	(0.048)	.0072	(0.048)
SX+	-.0914***	(0.026)	-.0811***	(0.026)	-.0512	(0.044)	-.0276	(0.043)
Age & gender			X		X		X	
Gay×gender					X		X	
Test Scores							X	

Note: *** p<0.01, ** p<0.05, * p<0.1. The White-collar variable takes the value of 1 if person's occupation is coded between 1 and 4 by the one-digit ISCO88 codes, and zero otherwise. Age enters in a quadratic polynomial. The estimations with test scores include six scores, one related to school grades (Math, Norwegian and English), two of them relate to agreeableness and the ease to make friends, the others relate to positivism, leadership and scholastic competence. All of them collected in the first wave of the survey.

the results. Second, it collects scores on personality traits and academic performance while kids. I use these measures to identify the unobserved skills, a key feature of my empirical strategy. Third, it collects detailed information about sexual preferences and practices. This way I do not run into absurdly small sample sizes and I am able to include single people in the analysis. And fourth, it collects information about the job market including ISCO88 coded occupations and yearly income.

YiN has four waves of data. The first wave was collected in 1992, when 12,287 students of lower and upper secondary school pupils, from 67 schools in grades 7-12 (age 12-20) were interviewed. The sampling allowed for each grade to be represented equally, and the sampling unit was the school. Wave 2 was collected two years later interviewing 7,637 youngsters. Some of the respondents answered the questionnaire in school and others answered it through mail. The study was originally planned to end after wave 2. However, a spin-off of the study continued in 1999 but only for those who had answered the wave 2 questionnaire in school and not by mail. 2,923 people were interviewed. Wave 4 was collected between May 2005 and August 2006. As in wave 3, the intended population to be interviewed were those who had answered the wave 2 questionnaire in school and not by mail. In this wave, data for 2,890 people were collected. There is a great overlap between those wave 3 and wave 4 respondents, in total 2,562 responded both surveys.

Following advise from the experts at NOVA, the agency responsible of the data, I do not include in my analysis respondents that were not followed-up in waves 3 or 4. This leaves a final sample of 2,991. According to Wichstrøm et al. (2013) those who attrited were more likely to be younger males with low parental SES, all of which are observables I can control for at every stage of the estimation.

3.3.1 Skills

The empirical strategy used in this Chapter—and presented in Section 3.4—relies on the identification of unobserved skills. Such identification and the subsequent estimations based on this unobserved heterogeneity pose heavy requirements regarding the minimum sample size. This minimum required increases geometrically with the number of unobserved dimensions of skill to be estimated. Given that the sample size of homosexuals in every definition is not large, I estimate the models with one unobserved factor. This factor will comprise both cognitive and non-cognitive skills in one construct.

To identify the unobserved skills factor, I use six scores; all of them collected in the first wave of the survey, when the respondents were in their early to mid teens. First, two of the measures relate to agreeableness, one in the area of social acceptance, and another one measuring how easy is for the respondent to build close friendships. The third measure relates to self-esteem. The fourth score relies on a scale that measures qualities like leadership and strong-mindedness. The fifth scale uses school grades in Math, Norwegian and English. Finally the sixth score measures scholastic competence.

3.3.2 Income Measures

The YiN data set does not contain a measure of hourly wage, instead it asks about pre-tax personal yearly income. Given that I am interested in discrimination in the work place, I want to focus on samples whose economic activity ensures that most of their income comes from their work. Namely, those who are employed ("Employed", 1961 observations), and those who report being employed full time ("Full Time", 1529 observations).[13] This way, I exclude inactive or unemployed population whose income may come from rehabilitation, disability or unemployment benefits.[14]

3.4 *Empirical Strategy*

The empirical strategy used in this Chapter corresponds to the one explained in detail in Section 1.5. Here, I will provide a short summary.

3.4.1 *The Difference Between Scores and Skills*

Works like Murnane et al. (1995); Heckman et al. (2006a); Espinoza et al. (2014) have shown that skills are relevant determinants of life choices and adult outcomes, like schooling, earnings and health, just as other characteristics like gender or household composition. Hence, the outcomes $\mathbf{Y}$ we are interested in (e.g., ln wage, schooling) are a result of observable characteristics $\mathbf{X_Y}$ (e.g., gender, tenure, local labor market) and skills θ. That is,

$$\mathbf{Y} = \mathbf{X}_Y \beta^Y + \alpha^Y \theta + \mathbf{e^Y} \tag{3.2}$$

where $\mathbf{Y}$ is a $M \times 1$ vector of outcome variables, $\mathbf{X}_Y$ is a matrix with all observable controls for each outcome variable, and $\mathbf{e^Y} \perp (\theta, \mathbf{X_Y})$ is a vector of error terms

[13] This sample selection, needed to be able address wage discrimination, could be troublesome as the observations left out of the sample may have characteristics that could influence in the final results. However, and to the advantage of the present work, if anything, the effect of the bias would make me results about a lower bound. That is because homosexuals are between 3 and 9 percentage points *less* likely to be employed (i.e., appear in the selected sample). Hence, if the unobservables that make a person participate in the labor force are positively correlated with income, then the homosexuals I get in the sample would have "better" unobservables than the ones left out, which would cause the estimated mean incomes of homosexuals to be slightly more than what the latent income is. Therefore, the estimated income gaps against homosexual would be smaller than if I had an uncensored sample.

[14] An important stylized fact is that homosexuals are equally likely to receive social benefits as heterosexuals are, regardless of the definition of homosexuality. Therefore, there is no reason to believe that any leakage in the income measures constructed could bias the results in either direction.

with distributions $f_{e^{y_m}}(\cdot)$ such that $e^{y_i} \perp e^{y_j}$ for every $m = 1, \ldots, M$. However, a difficulty arises in estimating equation (3.2): skills are not well defined entities with established scales or units of measurement. Instead they are latent variables that need to be inferred from variation captured in manifest variables they affect (Bartholomew et al., 2011). Hence, these manifest variables are assumed to be the result of a linear production function of test scores whose inputs are both observable characteristics and skills of the form:

$$\mathbf{T} = \mathbf{X}_T \beta^T + \alpha^{\mathbf{T}} \theta + \mathbf{e}^{\mathbf{T}} \tag{3.3}$$

where $\mathbf{T}$ is a $L \times 1$ vector of measurements (e.g., test scores), $\mathbf{X}_T$ is a matrix with all observable controls that affect each score, and $\mathbf{e}^{\mathbf{T}} \perp (\theta, \mathbf{X}_T)$ is a $L \times 1$ vector of mutually independent error terms that have associated distributions $f_{e^h}(\cdot)$ for every $h = 1, \ldots, L$. Being the skills latent, the best we can do as econometricians is to infer their underlying distribution $f_\theta(\cdot)$. That is the purpose of measurement system (3.3): use variation observed in the manifest variables to identify the distributional parameters of the unobserved factor (Kotlarski, 1967).[15] This facilitates the estimation the complete structural model because such distributions allow us to integrate away the unobservable variation of skill endowments in all the outcomes, choices and scores associated with the model, while still being able to retrieve the coefficients related with the skills in every equation.

The complete structure of the model, including the parameters that describe the distributions of the underlying factor $f_\theta(\cdot)$, is estimated using maximum likelihood

[15] See full set of restrictions needed for identification of the loadings and the diagonal matrix of the variances of the factors Σ_θ in Section 1.5

estimation (MLE).[16] The likelihood function will be the following

$$\mathcal{L} = \prod_{i=1}^{N} \int \left[\begin{array}{c} f_{e^{y_1}} \left(\mathbf{X}_{Y_1}, Y_1, \zeta \right) \times \cdots \times f_{e^{y_M}} \left(\mathbf{X}_{Y_M}, Y_M, \zeta \right) \\ \times f_{e^1} \left(\mathbf{X}_{T_1}, T_1, \zeta \right) \times \cdots \times f_{e^L} \left(\mathbf{X}_{T_L}, T_L, \zeta \right) \end{array} \right] dF_\theta \left(\zeta \right) \tag{3.4}$$

3.4.2 Roy Model: Allowing for Endogenous Choice

A model of potential outcomes inspired by the Roy model (Roy, 1951; Willis and Rosen, 1979) can easily be seen as a especial case of the setup presented above in subsection 3.4.1. For that, suppose $M = 3$ in the dimensions of vector $\mathbf{Y}$ in (3.2), where one row contains a binary choice D (e.g., going to college or not) and the other two contain the outcomes of interest that depend on such choice (e.g., wage for those that went to college and wage for those that did not). Hence, (3.2) becomes

$$\mathbf{Y} = \begin{bmatrix} D \\ Y_1 \\ Y_0 \end{bmatrix} = \begin{bmatrix} \mathbb{1} \left[\mathbf{X}_D \beta^{Y_D} + \alpha^{Y_D} \theta + e^D > 0 \right] \\ D \left(\mathbf{X}_Y \beta^{Y_1} + \alpha^{Y_1} \theta + e^{Y_1} \right) \\ (1 - D) \left(\mathbf{X}_Y \beta^{Y_0} + \alpha^{Y_0} \theta + e^{Y_0} \right) \end{bmatrix} \tag{3.5}$$

where $\mathbb{1}[A]$ denotes an indicator function that takes a value of 1 if A is true, and $\mathbf{X}_D$ represents a set of exogenous observables that affect the choice. In order for the model not to rely on its non-linearity for coefficient identification $\mathbf{X}_D$ needs to have a source of variation that is not present in $\mathbf{X}_Y$ and that affects Y_1 and Y_0 only through D.

An important feature of the Roy Model is that individuals must choose between

[16] As in the case of Chapter 1, I use the `heterofactor` command for Stata to perform the structural estimations presented in this Chapter (Sarzosa and Urzua, 2012).

two sectors, for example, treated and not treated. Therefore, both the decision and the outcomes are endogenous based on observable and unobservable characteristics of the individuals. This feature is key in the development of counterfactuals in Section 3.5.6. This empirical strategy is able to disentangle all the endogenous variables because it assumes that after controlling for the unobserved heterogeneity θ, the remaining error terms are independent (i.e., $e^D \perp e^{Y_1} \perp e^{Y_0}$), and can be modeled through independent contributions to likelihood function. That, is:

$$\mathcal{L} = \prod_{i=1}^{N} \int \begin{bmatrix} f^D\left(\mathbf{X}_D, Y_D, \zeta\right) f_{e^{Y_0}}\left(\mathbf{X}_Y, Y_0, \zeta\right)^{1-D} f_{e^{Y_1}}\left(\mathbf{X}_Y, Y_1, \zeta\right)^D \\ f_{e^1}\left(\mathbf{X}_{T_1}, T_1, \zeta\right) f_{e^2}\left(\mathbf{X}_{T_2}, T_2, \zeta\right) \times \cdots \times f_{e^L}\left(\mathbf{X}_{T_L}, T_L, \zeta\right) \end{bmatrix} dF_\theta\left(\zeta\right) \quad (3.6)$$

3.5 Results

This Section presents the main results of the Chapter of the dissertation. It is divided into several parts. First, I present the results of a structural model in which the unobserved skills are not estimated separately for homosexuals and heterosexuals. Instead, I estimate a national distribution of skills and capture the sexual minorities' income gap using dummy variables for homosexuals. This is the closest specification to that of the reduced-form results of the existing literature. Then, I move onto introduce this analysis into Roy models of potential outcomes. In the second part of this Section, I estimate separate skill distributions for homosexuals and heterosexuals and use them to estimate separate models. Then, I use simulations to show the existence and main drivers of the income gaps.

3.5.1 General Distribution of Skills: Income Gap Controlling for Unobservables

Let me first introduce the unobserved skills in a labor market outcome model similar to those of the existing literature. The idea is to expand the reduced-form results to incorporate unobserved heterogeneity. Hence the equation I estimate is

$$\ln Y = \alpha + \gamma Gay + \beta \mathbf{X} + \lambda \theta + \varepsilon$$

where θ is the unobserved skills that are drawn from a general distribution estimated using the complete sample. Like with the reduced-form results, I am interested in the estimates of γ. These are presented in Table 3.5 for each definition of homosexuality and each measure of yearly income. These results are similar to the reduced form ones from Table 3.2. They show evidence of income discrimination against homosexuals of about 10%, especially among the wider definitions of homosexuality.

Tab. 3.5: Ln Annual Income Gap Controlling for Skills

	SI	SP	SX	SX+
Employed	-0.113**	-0.107**	-0.057	-0.044
	(0.049)	(0.051)	(0.059)	(0.052)
Full Time	-0.107**	-0.097**	-0.058	-0.029
	(0.045)	(0.047)	(0.056)	(0.048)

Note: Standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1. All estimations were drawn from different structural estimations of income. The controls used and not presented in the Table are: Age, Age squared, gender, dummies for Father SES, population of the municipality and the unobservable factor of skills.

3.5.2 General Distribution of Skills: Roy Models

As explained above, in this subsection I introduce the unobserved heterogeneity and the Roy model of potential outcomes in a context where one general distribution of skills is estimated. Putting it in terms of (3.5), I estimate the following system

$$\mathbf{Y} = \begin{bmatrix} D \\ Y_1 \\ Y_0 \end{bmatrix} = \begin{bmatrix} \mathbb{1}\left[\gamma_D Gay + \mathbf{X}_D \beta^{Y_D} + \alpha^{Y_D}\theta + e^D > 0\right] \\ D\left(\gamma_{Y_1} Gay + \mathbf{X}_Y \beta^{Y_1} + \alpha^{Y_1}\theta + e^{Y_1}\right) \\ (1-D)\left(\gamma_{Y_0} Gay + \mathbf{X}_Y \beta^{Y_0} + \alpha^{Y_0}\theta + e^{Y_0}\right) \end{bmatrix}$$

where θ comes from a distribution estimated using the complete sample. In this case, the parameters of interest are γ_D, γ_{Y_0} and γ_{Y_1}. The advantage of this kind of models is that I am able to see differential effects of homosexuality depending whether $D = 1$ or $D = 0$, and allow the occupational and schooling choices to depend on characteristics of the respondent. I estimate two of these models that differ on the choice D made by the individual. In one model, I analyze the endogenous choice of schooling level. With it, I am able to inquire if there are different levels of discrimination depending on the schooling level. In the other one, I analyze the occupational choice. In this case I am able to see if there is a different type of discrimination agains homosexuals among white collar or blue collar occupations.

Table 3.6 presents the results of the Roy model of potential outcomes where the endogenous choice is tertiary education. The results show that, when controlling for skills, homosexuals are less likely to go to college, especially according to the narrowest definitions of homosexuality. Table 3.6 also shows that once in the labor market, homosexuals that undertook tertiary education face a significant wage gap, unlike those that did not go to college, who do not face any income gaps of their own.

The gaps faced by the college educated homosexuals range between 9% and 13% for those employed, and between 13% and 15% for those in full time jobs.

Table 3.7 on the other hand presents the results for a Roy model where the endogenous choice is whether to be a white collar or a blue collar worker. In congruence with the initial results of Table 3.4, the first row of Table 3.7 shows that after controlling for skills homosexuals tend to prefer blue collar occupations over white collar occupations. Interestingly, income discrimination against them is more pervasive in the former than in the latter. In fact, the income gap ranges from 8% to 24%, depending on the definition of homosexuality and yearly income used.

The results of Tables 3.6 and 3.7 attest to the importance of a model that allows for endogenous choices and separate estimates for two different sectors. This way, I find particular sectors where discriminations is stronger, namely among higher educated and among blue-collar occupations.

3.5.3 Separate Distributions: Unobserved Abilities

The main drawback of the empirical strategy used above is that it restricts the estimates of the model to be the same for homosexuals and heterosexuals. If there are any differences in the skill distributions, or in the labor market returns to observable and unobservable characteristics, the results presented above could be misleading. Therefore, from this point on, I present the results of models that allow to estimate a complete set of different parameters for heterosexuals and homosexuals, including the skills' distributions.

Figures 3.1 presents the estimated skill distributions for heterosexuals and homosexuals according to each definition of homosexuality. It is easy to see that not

Tab. 3.6: Effect of Homosexuality on ln Annual Income and College, Roy Model With Overall Distribution

	SI		SP		SX		SX+	
	Coeff.	Std.Err.	Coeff.	Std.Err.	Coeff.	Std.Err.	Coeff.	Std.Err.
College†	-0.046	(0.180)	-0.043	(0.102)	-0.523**	(0.242)	-0.627***	(0.210)
ln Income if Employed								
College	-0.087*	(0.046)	-0.137***	(0.048)	-0.135**	(0.058)	-0.103**	(0.051)
No College	-0.039	(0.048)	-0.001	(0.049)	0.008	(0.057)	-0.014	(0.047)
ln Income if Full Time								
College	-0.071	(0.048)	-0.137***	(0.052)	-0.131**	(0.060)	-0.148***	(0.054)
No College	-0.066	(0.050)	0.016	(0.049)	-0.020	(0.056)	0.010	(0.047)

Note: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. The coefficients presented are the ones associated with the homosexuality indicators. The College variable takes the value of 1 when the person has undertaken any kind of tertiary education in colleges (*høyskole*) or universities (*universitetet*). This includes three-year, four-year college studies or studies in the university with the same length. The controls used in the College choice equation and not presented in the Table are: age, age squared, gender, dummies for father's education, dummies for mother's education, whether as a child the person lived with one parent, a remarried parent or with no parents, number of books available at home as a child, and the unobservable factor of skills.. The controls used in the income equations and not presented in the Table are age, age squared, gender, dummies for Father SES, population of the municipality, height and the unobservable factor of skills.
† Coefficients, not marginal effects.

Tab. 3.7: Effect of Homosexuality on ln Annual Income and Occupation, Roy Model With Overall Distribution

	SI		SP		SX		SX+	
	Coeff.	Std.Err.	Coeff.	Std.Err.	Coeff.	Std.Err.	Coeff.	Std.Err.
White Collar	-.133	(0.097)	-.111	(0.1)	-.19	(0.12)	-.22**	(0.102)
Employed								
Blue Collar	-.238***	(0.052)	-.166***	(0.056)	-.129**	(0.063)	-.086*	(0.052)
White Collar	-.055	(0.046)	-.062	(0.049)	-.051	(0.059)	-.039	(0.051)
Full Time								
Blue Collar	-.245***	(0.064)	-.127**	(0.059)	-.113	(0.07)	-.08	(0.058)
White Collar	-.053	(0.063)	-.09	(0.068)	-.068	(0.077)	-.065	(0.069)

Note: *** p<0.01, ** p<0.05, * p<0.1. The coefficients presented are the ones associated with the homosexuality indicators. The White-collar variable takes the value of 1 if person's occupation is coded between 1 and 4 by the one-digit ISCO88 codes, and zero otherwise. The controls used in the College choice equation and not presented in the Table are: age, age squared, gender, dummies for father's education, dummies for mother's education, whether as a child the person lived with one parent, a remarried parent or with no parents, number of books available at home as a child, and the unobservable factor of skills.. The controls used in the income equations and not presented in the Table are age, age squared, gender, dummies for Father SES, population of the municipality, height and the unobservable factor of skills.

† Coefficients, not marginal effects.

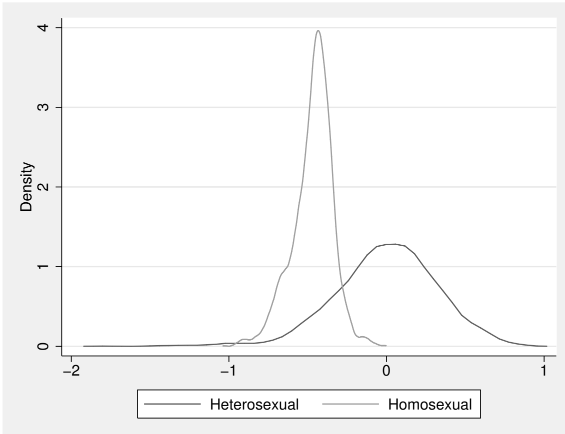
assuming normality and instead empirically estimating the distributions pays off as they are far from normally distributed, especially the ones from homosexuals. Figures 3.1 show that skills among homosexuals are more tightly distributed. While the standard deviation of skills among heterosexuals is around 0.33, it is between 0.12 to 0.22 for homosexuals. This interesting difference in the distribution of unobserved skills may have important consequences on schooling and occupational sorting and also in the income obtained when adults.

By construction, all the estimated distributions of unobserved ability are centered at zero.[17] However, Urzua (2008) showed that with a minor linearity assumption, one can use the difference between the estimated constants of the test score equations that contain the nummeraire loadings to infer the difference in means of the skill distributions of two groups. Figures 3.1 are depicted in a way such that the centers of the skills distributions of homosexuals are shifted to represent the difference in means with respect to those of heterosexuals. Importantly, none of these differences in the means of skill distributions between homosexuals and heterosexuals are statistically significant.

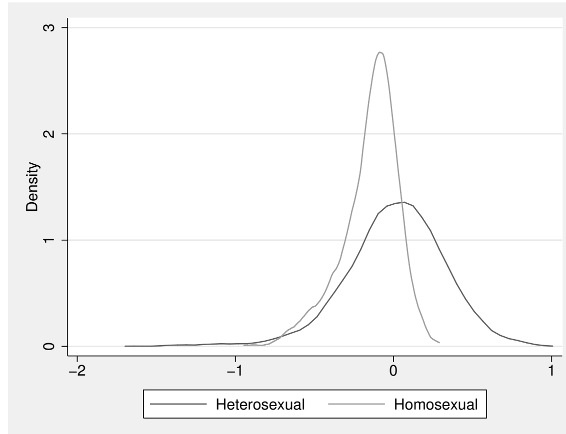
[17] In general, to estimate the adjunct measurement system (3.3), the empirical strategy uses Q normals that will be combined using weights ω_q. Abusing notation, the estimated distribution $f_\theta(\cdot) = \sum_{q=1}^{Q} \omega_q \mathcal{N}(\mu_q, \sigma_q)$, where $\sum_{q=1}^{Q} \omega_q = 1$. Hence, the weight of the last normal needs to fulfill $\omega_Q = 1 - \sum_{q=1}^{Q-1} \omega_q$. In the same way, given that all equations contain intercepts, the mean of the last normal μ_Q is one such that $\sum_{q=1}^{Q} \omega_q \mu_q = 0$. That is, $\mu_Q = \sum_{q=1}^{Q-1} \omega_q \mu_q / 1 - \sum_{q=1}^{Q-1} \omega_q$.

Fig. 3.1: Distribution of Unobserved Abilities by Sexual Preference

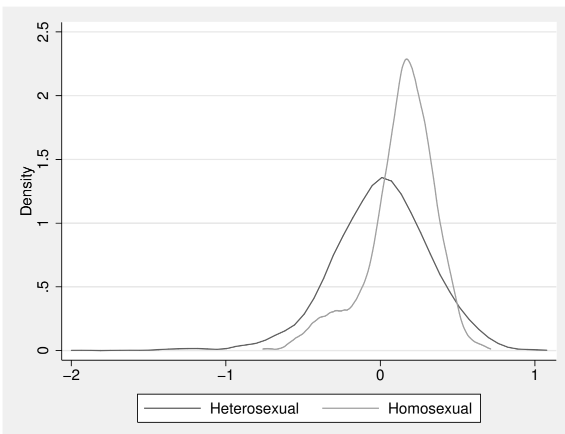
(a) According to SX Definition



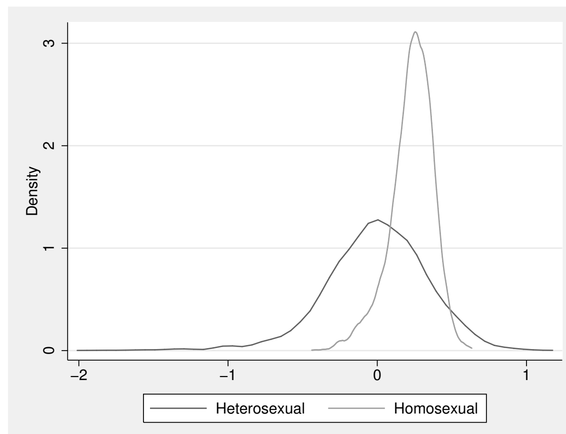
(b) According to SX+ Definition



(c) According to SP Definition



(d) According to SI Definition



3.5.4 Effect of Skills on Outcomes

With different estimated distributions of skills for every population group, I am able to estimate different models of potential outcomes and endogenous choices for each group. Therefore, the models I am estimating are the following

$$
\begin{bmatrix} D \\ Y_1 \\ Y_0 \end{bmatrix} = \begin{bmatrix} \mathbb{1}\left[\mathbf{X}_D \beta^{Y_D^S} + \alpha^{Y_D^S}\theta_S + e^{D,S} > 0\right] \\ D\left(\mathbf{X}_Y \beta^{Y_1^S} + \alpha^{Y_1^S}\theta_S + e^{Y_1^S}\right) \\ (1-D)\left(\mathbf{X}_Y \beta^{Y_0^S} + \alpha^{Y_0^S}\theta_S + e^{Y_0^S}\right) \end{bmatrix}
$$

$$
\begin{bmatrix} D \\ Y_1 \\ Y_0 \end{bmatrix} = \begin{bmatrix} \mathbb{1}\left[\mathbf{X}_D \beta^{Y_D^G} + \alpha^{Y_D^G}\theta_G + e^{D,G} > 0\right] \\ D\left(\mathbf{X}_Y \beta^{Y_1^G} + \alpha^{Y_1^G}\theta_G + e^{Y_1^G}\right) \\ (1-D)\left(\mathbf{X}_Y \beta^{Y_0^G} + \alpha^{Y_0^G}\theta_G + e^{Y_0^G}\right) \end{bmatrix} \tag{3.7}
$$

where the parameters labeled with S correspond to heterosexuals and the ones labeled with G correspond to homosexuals. Tables 3.8 and 3.9 present the standardized estimates of the sexual preference-specific coefficients associated with skills for income, tertiary education and occupation.[18] Table 3.8 shows that skills are important determinants of sorting into tertiary education regardless of the sexual orientation. However, the results also show that the returns to skills do differ between heterosexuals and homosexuals, supporting the use of a model that allows different parameters across sexual preferences. Interestingly, among heterosexuals, skills have significant positive returns only for those that undertook some sort of tertiary education. On the other hand, among homosexuals, skills are rewarded independently of the level of education attained. However, skills are rewarded more among those who went to tertiary education than among those who did not. The numbers in Table 3.8 also indicate that skills have stronger effects on income for homosexuals than for heterosexuals. In particular, while a one standard deviation increase in skills among college educated

[18] All the tables in this Subsection lack results for the SX definition of homosexuality because the models like (3.7) did not converge, due to insufficient sample size.

heterosexuals causes yearly income to increase about 10% to 25%, the yearly income increase due to a one standard deviation increase in skills among college educated homosexuals rages between 45% and 75%.

Table 3.9, in turn, shows that even though skills impact positively the likelihood of choosing a white collar occupation, in general blue collar occupations reward more those skills in terms of annual income. The numbers do not allow to specify whether skills are rewarded more in general among homosexuals or heterosexuals. None of the groups have an overall advantage over the other. However, there are specific definitions, like SI, where there seems to be a higher reward for skills among blue collar homosexuals than among heterosexuals in the same occupations.

Tab. 3.8: Standardized Loadings: ln Annual Income and Tertiary Education Choice

	Heterosexuals			Homosexuals		
	SI	SP	SX+	SI	SP	SX+
College	0.796***	0.811***	1.022***	0.349***	0.074	0.152*
	(0.19)	(0.157)	(0.191)	(0.082)	(0.11)	(0.087)
Employed						
No College	0.051	0.041	0.053	0.454***	0.395***	0.347***
	(0.037)	(0.037)	(0.05)	(0.011)	(0.011)	(0.007)
College	0.099*	0.065	0.160***	0.499***	0.617***	0.635***
	(0.052)	(0.049)	(0.058)	(0.031)	(0.046)	(0.014)
Full Time						
No College	0.023	0.024	0.026	0.403***	0.326***	0.275**
	(0.034)	(0.038)	(0.04)	(0.009)	(0.038)	(0.115)
College	0.07	0.045	0.045	0.524***	0.752***	0.454***
	(0.063)	(0.058)	(0.058)	(0.016)	(0.023)	(0.035)

Note: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parenthesis. The coefficients presented are the estimates of the effect of the unobserved factor on each equation of model (3.7). Models are estimates allowing for different distributions of the unobserved factor depending on sexual preference. The College variable takes the value of 1 when the person has undertaken any kind of tertiary education in colleges (*høyskole*) or universities (*universitetet*). This includes three-year, four-year college studies or studies in the university with the same length. The observable controls used in the College choice equation are: age, age squared, gender, dummies for father's education, dummies for mother's education, whether as a child the person lived with one parent, a remarried parent or with no parents, number of books available at home as a child. The observable controls used in the income equations are age, age squared, gender, dummies for Father SES, population of the municipality, height and the unobservable factor of skills.

Tab. 3.9: Loadings: ln Annual Income and Occupational Choice

	Heterosexuals			Homosexuals		
	SI	SP	SX+	SI	SP	SX+
White Collar	0.871***	0.752***	0.845***	1.546	2.355*	1.834
	(0.148)	(0.129)	(0.146)	(1.637)	(1.338)	(2.198)
Employed						
Blue Collar	0.091**	0.029	0.056	0.584***	0.453***	0.433***
	(0.045)	(0.037)	(0.047)	(0.024)	(0.054)	(0.035)
White Collar	0.006	-0.004	0.018	0.042	0.166	0.391*
	(0.05)	(0.049)	(0.05)	(0.194)	(0.171)	(0.206)
Full Time						
Blue Collar	0.033	0.053	0.041	0.01	0.027	-0.023
	(0.028)	(0.047)	(0.047)	(0.215)	(0.133)	(0.086)
White Collar	-0.031	-0.023	0.006	0.163	0.211	0.533**
	(0.05)	(0.054)	(0.056)	(0.142)	(0.212)	(0.216)

Note: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Standard errors in parenthesis. The coefficients presented are the estimates of the effect of the unobserved factor on each equation of model (3.7). Models are estimates allowing for different distributions of the unobserved factor depending on sexual preference. The White-collar variable takes the value of 1 if person's occupation is coded between 1 and 4 by the one-digit ISCO88 codes, and zero otherwise. The observable controls used in the occupation choice equation are: age, age squared, gender, dummies for father's education, dummies for mother's education, whether as a child the person lived with one parent, a remarried parent or with no parents, number of books available at home as a child. The observable controls used in the income equations are age, age squared, gender, dummies for Father SES, population of the municipality, height and the unobservable factor of skills.

3.5.5 *Goodness of Fit of the Model*

Tables 3.10 and 3.11 present evidence on the models' goodness-of-fit for (ln) annual income for each education level and occupational choice, respectively. The models do very well in predicting the means for each definition of income and homosexuality. The performance of the models predicting (ln) income for homosexuals is remarkably good considering the reduced sample size the models are facing.

Table 3.12 shows the models' goodness-of-fit for replicating the endogenous choices made by the respondents, namely college going and whether to be a white collar or a blue collar worker. Again, a comparison between the predicted means and the actual data shows that the models do very well in both choices and for every definition of homosexuality.

These results show that the models predict very well the labor market outcomes and choices for both homosexuals and heterosexuals. This provides confidence about the fact that the model will produce appropriate counterfactuals to successfully analyze inequality in the labor market due to sexual preferences.

Tab. 3.10: Goodness of Fit - Ln Annual Income by Schooling level and Sexual Preference

	Heterosexual			Homosexual		
	SI	SP	SX+	SI	SP	SX+
lnIncome if Employed						
No College						
Actual	5.556	5.558	5.563	5.442	5.453	5.446
Model	5.554	5.562	5.555	5.407	5.461	5.464
College						
Actual	5.514	5.525	5.512	5.432	5.354	5.398
Model	5.484	5.518	5.474	5.428	5.316	5.315
lnIncome if Full Time						
No College						
Actual	5.669	5.655	5.661	5.536	5.597	5.571
Model	5.64	5.625	5.634	5.529	5.542	5.539
College						
Actual	5.575	5.583	5.581	5.483	5.414	5.539
Model	5.571	5.578	5.573	5.491	5.322	5.503

Tab. 3.11: Goodness of Fit - Ln Annual Income by Occupational Choice and Sexual Preference

	Heterosexual			Homosexual		
	SI	SP	SX+	SI	SP	SX+
lnIncome if Employed						
Blue Collar						
Actual	5.451	5.449	5.429	5.205	5.212	5.319
Model	5.448	5.441	5.425	5.193	5.227	5.356
White Collar						
Actual	5.612	5.596	5.596	5.537	5.522	5.511
Model	5.631	5.619	5.616	5.551	5.526	5.512
lnIncome if Full Time						
Blue Collar						
Actual	5.613	5.577	5.567	5.329	5.397	5.448
Model	5.585	5.536	5.525	5.33	5.385	5.442
White Collar						
Actual	5.646	5.633	5.627	5.601	5.505	5.536
Model	5.662	5.655	5.646	5.605	5.509	5.531

Tab. 3.12: Goodness of Fit - Schooling level and Occupation by Sexual Preference

	Heterosexual			Homosexual		
	SI	SP	SX+	SI	SP	SX+
Prob. College Going						
Actual	0.627	0.641	0.655	0.705	0.644	0.570
Model	0.633	0.641	0.655	0.691	0.636	0.561
Prob. White Collar						
Actual	0.598	0.607	0.614	0.612	0.564	0.526
Model	0.598	0.606	0.611	0.595	0.551	0.525

3.5.6 Gaps in Labor Market Outcomes

From the results I have provided so far, it is possible to argue three facts. First, that unobserved skills are important determinants of education choices, occupational choices and income earned. Second, that the way these skills affect those adult outcomes depends on the sexual preference of the individual. And third, that skills follow different distributions for homosexuals and heterosexuals. Hence, based on these three facts, one could hypothesize that the income gaps observed could respond to these estimated difference in skill distributions, and that the amount of the gap that is not explained by such differences—or differences in observable characteristics—could be attributed to plain discrimination.

Urzua (2008) showed that one can use the structure of the model to evaluate these claims. In particular, it is possible to use the model to simulate the outcomes that homosexuals would have obtained if their skills came from the heterosexual skills' distribution and/or if they had the observable characteristics homosexuals have.[19] Following that approach and using the same notation as in equation (3.7), let

$$Y^{H}\left(\mathbf{X}_{Y}^{H_X},\mathbf{X}_{D}^{H_X},\theta^{H_\theta}\right)=D^{H}\left(\mathbf{X}_{D}^{H_X},\theta^{H_\theta}\right)Y_1^{H}\left(\mathbf{X}_{Y}^{H_X},\theta^{H_\theta}\right)$$
$$+\left[1-D^{H}\left(\mathbf{X}_{D}^{H_X},\theta^{H_\theta}\right)\right]Y_0^{H}\left(\mathbf{X}_{Y}^{H_X},\theta^{H_\theta}\right)$$

be the (ln) income given by observable characteristics $\mathbf{X}_{Y}^{H_X}$ and $\mathbf{X}_{D}^{H_X}$ and unobservable skills θ^{H_θ}. The supra-index denotes the sexual preference group (i.e., $H_{\bullet}=\{S,G\}$). Hence, for instance, income for heterosexuals is given by $Y^{S}\left(\mathbf{X}_{Y}^{S},\mathbf{X}_{D}^{S},\theta^{S}\right)$, while income for homosexuals is given by $Y^{G}\left(\mathbf{X}_{Y}^{G},\mathbf{X}_{D}^{G},\theta^{G}\right)$. These are the values presented in Tables 3.10 and 3.11. I can use this notation to identify the simulation of counterfactuals. For instance, $Y^{G}\left(\mathbf{X}_{Y}^{G},\mathbf{X}_{D}^{G},\theta^{S}\right)$ denotes the simulated income earned by homosexuals when given skills drawn from the heterosexuals' skills distributions but keeping the observable characteristics as they are. Consequently, the simulated income earned by homosexuals when given the skills and observable characteristics of

[19] See the differences in observable characteristics across groups in Table N.1 in the Appendix

heterosexuals is denoted by $Y^G\left(\mathbf{X}_Y^S, \mathbf{X}_D^S, \theta^S\right)$.

The fact that choices $D^H\left(\mathbf{X}_D^{H_X}, \theta^{H_\theta}\right)$ are endogenous is an important feature of the models, that affects the simulation of counterfactuals. That is, when counterfactuals are simulated, their new results are product not only of direct changes in the outcome equations $Y_1^H\left(\mathbf{X}_Y^{H_X}, \theta^{H_\theta}\right)$ and $Y_0^H\left(\mathbf{X}_Y^{H_X}, \theta^{H_\theta}\right)$, but also of indirect changes through the schooling or occupational decision. Hence, for instance, when simulating the model for homosexuals but giving them skills drawn from the heterosexuals's distribution, the choice becomes $D^G\left(\mathbf{X}_D^G, \theta^S\right)$. This materializes the very important fact that education and occupational decisions are taken based on observable and unobservable characteristics, and once these change, people might choose differently. Hence, the measure of the total income gaps between the two groups need to incorporate the differences that arise from both the direct and indirect channels. That is, the different gaps measured—depending on the counterfactual used—are provided by

$$\begin{aligned}
&E\left[Y^S\left(\mathbf{X}_Y^S, \mathbf{X}_D^S, \theta^S\right)\right] - E\left[Y^G\left(\mathbf{X}_Y^{H_X}, \mathbf{X}_D^{H_X}, \theta^{H_\theta}\right)\right] \\
&\qquad = \Pr\left[D^S\left(\mathbf{X}_D^S, \theta^S\right) = 1\right] E\left[Y_1^S\left(\mathbf{X}_Y^S, \theta^S\right) \middle| D^S\left(\mathbf{X}_D^S, \theta^S\right) = 1\right] \\
&\qquad + \Pr\left[D^S\left(\mathbf{X}_D^S, \theta^S\right) = 0\right] E\left[Y_0^S\left(\mathbf{X}_Y^S, \theta^S\right) \middle| D^S\left(\mathbf{X}_D^S, \theta^S\right) = 0\right] \\
&\qquad - \Pr\left[D^G\left(\mathbf{X}_D^{H_X}, \theta^{H_\theta}\right) = 1\right] E\left[Y_1^G\left(\mathbf{X}_Y^{H_X}, \theta^{H_\theta}\right) \middle| D^G\left(\mathbf{X}_D^{H_X}, \theta^{H_\theta}\right) = 1\right] \\
&\qquad - \Pr\left[D^G\left(\mathbf{X}_D^{H_X}, \theta^{H_\theta}\right) = 0\right] E\left[Y_0^G\left(\mathbf{X}_Y^{H_X}, \theta^{H_\theta}\right) \middle| D^G\left(\mathbf{X}_D^{H_X}, \theta^{H_\theta}\right) = 0\right] \quad (3.8)
\end{aligned}$$

Notice that differences in schooling or occupational choices across groups (i.e., $\Pr\left[D^S\left(\mathbf{X}_D^S, \theta^S\right) = 1\right]$ versus $\Pr\left[D^G\left(\mathbf{X}_D^{H_X}, \theta^{H_\theta}\right) = 1\right]$) are important in determining the overall income gap because they mediate the way income differences between homosexuals and heterosexuals for each level of education or occupational choice enter in the overall gap calculation.

Tables 3.13 and 3.15 present the results using several counterfactual simulations on the different definitions of yearly income and homosexuality, and Tables 3.14 and 3.16 present the results for the proportion of people taking-up tertiary education and white collar occupations respectively. In Tables 3.13 and 3.15, panels A and B present the results for homosexuals and heterosexuals with their own characteristics.[20] The following two panels present the calculated income gaps described by (3.8). The difference between the two panels is that in the one labeled *Gap (Equalizing Gender)* the simulations correct for the fact that there are more women than men defined as homosexuals (see Table 3.1) and if women earn consistently less than men, the gaps I calculate could be confounding the effect of belonging to a sexual minority with the effect of being a woman. By equalizing the gender distributions across groups, I prevent gender discrimination biasing the comparisons. Panels C present the results that arise after changes in skill distributions (i.e., Panel C1), observables (i.e., Panel C2), and both skills and observables (i.e., Panel C3).[21]

3.5.6.1 Model With Education Choice

Tables 3.13 and 3.14 present the simulated counterfactual results for the models with education choice. Before going into the analysis of the income gaps against homosexuals, let me first show an important feature of the Norwegian labor market: the fact that people with tertiary education do not earn more—on average—than people without this level of education. This can be seen by comparing the simulated incomes across educational choices. For instance, among employed heterosexuals people that

[20] Notice that the figures presented there are the same as the one presented in Table 3.10 and 3.11.

[21] Results in presented in Panel C1 also correct for gender disparities in definition of homosexuals.

did not undertake tertiary education earn around 7% more than college educated workers. This can be due to the strength of Norwegian practical training, or the fact that the people in the sample are still in their late 20s and tertiary education might become more valuable later in life.[22] The explanation of this phenomenon is not of interest for this work, but this feature of the Norwegian labor market it is key in explaining some of the results below, especially because I am endogenizing schooling choices and sorting into schooling may play a crucial role in explaining the income gaps against sexual minorities.

Panels B1 and B2 of the Table 3.13 present the income gaps for each sample and homosexuality definition without and with homogenized gender proportions, respectively. Notice that homogenizing gender proportions in general reduce the point estimates of the overall gap, although by a very small margin. The biggest changes after this correction are seen not in the overall gaps, but in the composition of the gaps for SI and SP homosexuals. The correction reduces the gaps estimated for those that did not go to college, while it increases them among those that did. These recomposition of the gaps respond to the fact that there is a different sorting into schooling, as can be seen in Table 3.14. The confounding effect that gender differences have misleadingly increase the difference in take-up of tertiary education between heterosexuals and homosexuals, which in turn affects the composition of the gaps between the ones faced by those that go to college and the ones face by those that did not.

[22] Figure N.1 in the Appendix shows that even though there is an income difference against the younger college educated workers that disappears after they are 27 years of age, the rate of growth of income of the college educated flattens and equalizes that of the non-college educated after they are 28. Hence, there is evidence against the hypothesis that college may be rewarded more among older workers.

Tab. 3.13: Heterosexual and Homosexual ln Annual Income Under Different Assumptions

	No College			College			Overall		
	SI	SP	SX+	SI	SP	SX+	SI	SP	SX+
A. Outcomes Heterosexuals									
Employed	5.554	5.562	5.555	5.484	5.518	5.474	5.51	5.534	5.502
Full Time	5.64	5.625	5.634	5.571	5.578	5.573	5.596	5.595	5.594
B1. Outcomes Homosexuals									
Employed	5.407	5.461	5.464	5.428	5.316	5.315	5.422	5.367	5.378
Full Time	5.529	5.542	5.539	5.491	5.322	5.503	5.502	5.4	5.519
Gap									
Employed	0.465	0.059	-0.419	-0.377	0.108	0.542	0.088	0.167	0.124
Full Time	0.433	0.01	-0.464	-0.339	0.184	0.539	0.094	0.195	0.075
B2. Outcomes Homosexuals (Equalizing Gender)									
Employed	5.437	5.511	5.535	5.42	5.31	5.333	5.426	5.384	5.418
Full Time	5.619	5.588	5.589	5.459	5.366	5.512	5.512	5.451	5.545
Gap (Equalizing Gender)									
Employed	0.279	-0.05	-0.433	-0.195	0.2	0.517	0.084	0.15	0.084
Full Time	0.169	-0.165	-0.51	-0.086	0.308	0.559	0.084	0.144	0.049
C1. Outcomes Homosexuals with Heterosexuals Unobservables									
Employed									
Income	5.05	5.432	5.384	5.662	5.393	5.55	5.466	5.407	5.481
Gap	0.41	0.038	-0.34	-0.365	0.089	0.361	0.044	0.127	0.021
Full Time									
Income	5.522	5.557	5.526	5.535	5.415	5.596	5.531	5.469	5.566
Gap	0.217	-0.126	-0.425	-0.152	0.252	0.453	0.065	0.126	0.028
C2. Outcomes Homosexuals with Heterosexuals Observables									
Employed									
Income	5.38	5.432	5.426	5.441	5.368	5.47	5.422	5.388	5.453
Gap	0.39	0.277	-0.135	-0.302	-0.131	0.183	0.088	0.146	0.049
Full Time									
Income	5.607	5.572	5.564	5.467	5.385	5.551	5.509	5.447	5.556
Gap	0.335	0.108	-0.208	-0.248	0.039	0.245	0.087	0.148	0.038

Continued on next page

Table 3.13 – continued from previous page

	No College			College			Overall		
	SI	SP	SX+	SI	SP	SX+	SI	SP	SX+
C3. Outcomes Homosexuals with Heterosexuals Observables & Unobservables									
Employed									
Income	4.941	5.385	5.295	5.677	5.404	5.628	5.441	5.398	5.5
Gap	0.445	0.27	-0.133	-0.376	-0.134	0.135	0.069	0.136	0.002
Full Time									
Income	5.459	5.557	5.517	5.552	5.405	5.600	5.524	5.456	5.568
Gap	0.369	0.107	-0.195	-0.297	0.031	0.221	0.072	0.139	0.026

The results in panel B of Table 3.13 show that the estimated income gaps faced by sexual minorities range from around 5% to 15%, depending on the sample and on the homosexuality definition used. In general, the largest gaps are estimated among the college educated workers. In fact, in the case of SP and SX+ definitions of homosexuality, the income differences among college educated workers contribute positively to the overall gap, while the income differences among those that did not go to college contribute negatively to the overall gap.[23] Although these two opposing effects counter each other, the gaps faced by college educated homosexuals outweigh the income advantage that homosexuals that did not go to college have. Interestingly, the income differences estimated under the SI homosexuality definition work differently. SI homosexuals face income differences of between 8.4% overall, and the main driver of those differences is the the gaps faced by those who did not attend tertiary education. This responds to the fact that while homosexuals under other definitions are always less likely to have tertiary education than heterosexuals, SI homosexuals are much more likely to do so than SI heterosexuals (see Table 3.14).

In general, the definition that reports the largest income difference is SP, for which the income differences between homosexuals and heterosexuals is of 15% in the employed sample and 14.4% among full time workers. On the other had, the tightest definition of homosexuality is the one that report the smallest income differences ranging between 5% and 8.4%.

[23] Even though both college educated and not college educated homosexuals face income gaps against them, the gap faced in each sector contributes in different directions because of the differences in college attendance between homosexuals and heterosexuals. Table 3.14 shows that SP and SX+ homosexuals are less likely to go to college than heterosexuals.

Tab. 3.14: College Attendance Under Different Assumptions

	SI	SP	SX+
A. Heterosexuals			
	0.6327	0.6411	0.6553
B.1 Homosexuals			
	0.6907	0.6365	0.5660
B.2 Homosexuals (Equalizing Gender)			
	0.6548	0.6151	0.5646
C.1 Homosexuals with Heterosexuals Unobservables			
	0.6521	0.6262	0.5705
C.2 Homosexuals with Heterosexuals Observables			
	0.6648	0.6699	0.61
C.3 Homosexuals with Heterosexuals Unobservables & Observables			
	0.647	0.6695	0.6087

Panel C.1 presents the simulation results where homosexuals have been given skills taken from the heterosexuals' skills distribution (i.e., $Y^G\left(\mathbf{X}_Y^G, \mathbf{X}_D^G, \theta^S\right)$). These results show that equalizing skills across groups reduces the income gaps against homosexuals. The extent of that reduction depends on the homosexuality definition used. There are larger reduction in the gap for SI and SX+ homosexuals than for SP homosexuals. For instance, the gap faced by the employed SI homosexuals falls from 8.4% to 4.4%, and those employed in full time occupations would see their income gap drop from 8.4% to 6.5%. The income difference against SX+ full time employed homosexuals would go down to by 2.1 percentage points (from 4.9% to 2.8%), and among the employed sample, it would be reduced to a fourth of what it is (from 8.4% to 2.1%). The main driver of all these reductions in the overall income differences is the increase in income among college educated homosexuals that would result from them having their skills drawn from the skills distribution of their heterosexual counterparts. Furthermore, panel C1 of Table 3.14 indicates that equalizing skill distributions does not affect average tertiary education take-up. Hence all the difference in gaps presented in Table 3.13 respond to changes in the differences in income and not difference in sorting.

Panel C.2 shows the contribution of observable characteristics in determining income differences. Panel C.2 shows that when homosexuals are given heterosexuals observable characteristics (i.e., $Y^G\left(\mathbf{X}_Y^S, \mathbf{X}_D^S, \theta^G\right)$), the income gaps remain mostly unchanged. Only the SX+ homosexuals would see their gaps fall in this scenario, and the gap reduction is smaller than what would happen if homosexuals were given skills taken from the heterosexuals' skills distribution. Different reasons explain the lack of change in the gaps that SI and SP homosexuals would face after receiving heterosexuals' observables. While in the case of SI homosexuals education-specific gaps and college take-up would remain unchanged, for SP homosexuals education-specific gaps and tertiary education take-up would be adjusted. In particular, providing SP homosexuals with heterosexual observable characteristics would cause a small income increase among those that did not go to college, while it would reduce the income earned by those who went to college. These effects would be amplified by 5.5 percentage points increase in the likelihood of going to college. This new sorting would cause all the gains and losses to cancel each other out leaving overall gaps almost unchanged.

Providing SX+ homosexuals with heterosexuals observable characteristics would cause an improvement of the conditions for the college educated ones. It would make

SX+ college educated homosexuals earn almost as much as heterosexuals. Although there is a deterioration income differences non-college educated SX+ homosexuals face, it would not be enough to offset the gains among the college educated ones. In addition, panel C.2 in Table 3.14 shows that SX+ homosexuals would also have a significant gain in the likelihood of going to college.

Finally, panel C.3 shows the counterfactuals when homosexuals are given both observables and unobservables of heterosexuals (i.e., $Y^G\left(\mathbf{X}_Y^S, \mathbf{X}_D^S, \theta^S\right)$). Not surprisingly, the changes in the gaps presented in panel C.3 are a combination of the ones presented in panels C1 and C2. In consequence, there would be a smaller reduction in the gaps in wider definitions of homosexuality than in the strictest definition of it. That is because while changing skill endowments of SI and SP homosexuals would reduce the gaps, changing the observables characteristics would not. On the other hand, both changes work in the same direction for the SX+ homosexuals. Therefore, joint changes of skill endowments and observable characteristics reach deeper income gaps reductions. So while these gaps would be now of the order of 7% for SI homosexuals and 13% for SP homosexuals, they would range between 0.2% and 2.6% for the SX+ ones.

3.5.6.2 Model With Occupational Choice

Now, I turn to the model with occupational choice. Table 3.15 shows the income differences that arise from different counterfactual simulation exercises. Notice that workers in white collar occupations earn significantly more than workers in blue collar occupations. This will be an important driver of the income gaps faced by members of sexual minorities because, as can be seen in Table 3.16, they tend to prefer white collar occupations significantly less than heterosexual workers. Therefore, any changes than boost homosexual take-up of white collar occupations will have a big impact in reducing the income gaps they face.

The results in Table 3.15 show that the existing overall income gaps range between 10% and 15% when taking into account occupational choices. This matches with the overall income gaps estimated by the model with education choice in Table 3.13.[24] As in the case of the models with education choices, these results show smaller gaps for the most strict definition of homosexuality. Although the income differences against members of sexual minorities within each type of occupation are not that different, their reduced participation in white collar occupations compared to heterosexuals (see Table 3.16) amplify the income gaps homosexuals face in this type of occupations, which end up contributing to the overall gap with 27% to 57%.

Panel C1 in Table 3.15 shows that equalizing the distributions of skills would make homosexuals in white collar occupations earn more, however it would also affect the sorting into this type of occupations making them even less likely to opt into them. Hence, the overall gaps increase for the wider definitions of homosexuality. For SX+ homosexuals, on the other hand, the effects of equalizing skill distributions differ depending on the sample. Among the employed population, the increase in income that homosexual white collar workers would enjoy would not be enough to reduce the overall gap because it is not enough to offset the effect of the fall in white collar occupations take-up by homosexuals. On the other hand, homosexual full time workers would see an income increase of 30% which is able to compensate the fact that less SX+ homosexuals are opting into white collar occupations. Hence for this

[24] Small discrepancies in the estimated overall gaps could have been expected as the two models use different samples. That is because 15% of the observations with positive income do not report their ISCO88 occupation code.

population, the overall income gaps would disappear.

Panel C2 shows the effect of equalizing observable characteristics. It shows that although the incomes generated in each type of occupation would not differ from the actual ones, giving homosexuals the observable characteristics of heterosexuals would boost sorting into white collar occupations, even surpassing the actual take up of heterosexuals. This new sorting into higher paying occupations would shift the overall income differences dramatically. In this scenario, homosexuals would not face income gaps agains them, instead they would have an advantage over heterosexual workers that would range from 33% to 50%.

Panel C3 Table 3.16 shows that, when homosexuals are given heterosexuals' observable and unobservable characteristics, there would also be a different sorting into occupations that would increase the proportion of homosexuals in white collar jobs. As it was shown in panel C2 of Table 3.15, this new sorting is due to changes in the observable characteristics. In consequence, under this scenario also, homosexuals would have an advantage over heterosexual workers in terms of income earned.

All the evidence presented in Table 3.15 indicates that occupational choice are key in determining income differences agains members of sexual minorities. Although they earn less regardless of the occupation, their preference for blue collar occupations exacerbate the income gaps. This preference is closely linked to observable characteristics, that once modified make more homosexuals sort into white collar occupations causing overall income gaps to disappear. Interestingly, giving unobserved skills to homosexuals form the skills distribution of heterosexuals would increase their income earned in white collar occupations, but would not modify the proportion of homosexuals preferring blue collar occupations.

Tab. 3.15: Heterosexual and Homosexual ln Annual Income Under Different Assumptions, By Occupation

	Blue Collar			White Collar			Overall		
	SI	SP	SX+	SI	SP	SX+	SI	SP	SX+
A. Outcomes Heterosexuals									
Employed	5.448	5.441	5.425	5.631	5.619	5.616	5.558	5.549	5.542
Full Time	5.585	5.536	5.525	5.662	5.655	5.646	5.631	5.608	5.599
B1. Outcomes Homosexuals									
Employed	5.193	5.227	5.356	5.551	5.526	5.512	5.405	5.391	5.438
Full Time	5.33	5.385	5.442	5.605	5.509	5.531	5.493	5.453	5.489
Gap									
Employed	0.062	-0.213	-0.424	0.09	0.371	0.528	0.153	0.158	0.104
Full Time	0.061	-0.247	-0.426	0.077	0.402	0.536	0.138	0.155	0.11
B2. Outcomes Homosexuals (Equalizing Gender)									
Employed	5.19	5.243	5.378	5.599	5.562	5.502	5.417	5.413	5.443
Full Time	5.350	5.399	5.471	5.622	5.49	5.507	5.501	5.447	5.49
Gap (Equalizing Gender)									
Employed	-0.13	-0.316	-0.46	0.271	0.452	0.559	0.141	0.136	0.099
Full Time	-0.146	-0.352	-0.465	0.276	0.512	0.574	0.13	0.161	0.109
C1. Outcomes Homosexuals with Heterosexuals Unobservables									
Employed									
Income	4.559	5.084	5.103	5.63	5.627	5.74	5.135	5.362	5.428
Gap	0.069	-0.342	-0.387	0.354	0.529	0.501	0.423	0.187	0.114
Full Time									
Income	5.337	5.397	5.476	5.772	5.575	5.855	5.569	5.488	5.671
Gap	-0.258	-0.452	-0.509	0.319	0.571	0.437	0.062	0.12	-0.072
C2. Outcomes Homosexuals with Heterosexuals Observables									
Employed									
Income	5.289	5.239	5.341	5.522	5.485	5.445	5.889	6.007	6.058
Gap	-0.269	-0.417	-0.504	-0.063	-0.04	-0.013	-0.331	-0.458	-0.516
Full Time									
Income	5.411	5.418	5.478	5.526	5.449	5.432	5.97	6.061	6.09
Gap	-0.292	-0.462	-0.51	-0.046	0.009	0.019	-0.339	-0.453	-0.491

Continued on next page

Table 3.15 – continued from previous page

	Blue Collar			White Collar			Overall		
	SI	SP	SX+	SI	SP	SX+	SI	SP	SX+
C3. Outcomes Homosexuals with Heterosexuals Observables & Unobservables									
Employed									
Income	4.719	5.164	5.168	5.549	5.509	5.55	5.56	5.908	5.973
Gap	-0.005	-0.381	-0.419	0.004	0.022	-0.013	-0.002	-0.359	-0.431
Full Time									
Income	5.401	5.412	5.488	5.628	5.481	5.58	5.95	6.001	6.121
Gap	-0.287	-0.459	-0.515	-0.031	0.066	-0.008	-0.319	-0.393	-0.522

Tab. 3.16: Occupation Under Different Assumptions

	SI	SP	SX+
A. Heterosexuals			
	0.6004	0.6066	0.6119
B.1 Homosexuals			
	0.5982	0.5525	0.525
B.2 Homosexuals (Equalizing Gender)			
	0.5569	0.5324	0.5200
C.1 Homosexuals with Heterosexuals Unobservables			
	0.5378	0.5122	0.5107
C.2 Homosexuals with Heterosexuals Observables			
	0.6236	0.6307	0.6322
C.3 Homosexuals with Heterosexuals Unobservables & Observables			
	0.6093	0.6151	0.6193

3.6 Conclusions

This chapter of the dissertation incorporates differences in skill distributions, observable characteristics and tastes for schooling and occupations into the analysis and quantification of the income gaps observed against non-heterosexual workers. The results from the model indicate that there are in fact differences in the the distribution of unobserved skills and observed characteristics between homosexual and heterosexual workers. These traits work different for sexual preference groups. That is, they have different effects on education and educational choices, as well as different return in terms of income. In particular, while skills have a stronger effect on sorting into tertiary education and white collar occupations for heterosexuals, skills are more rewarded in terms of income among homosexuals than among heterosexuals. In addition, observed characteristics are strong determinants of the difference in occupation choices observed between sexual preference groups. All these results indicate that there are several forces working simultaneously in opening the income gaps observed.

The findings of this Chapter indicate that, assuming that tastes for occupational choices remain unaltered, differences in skill distributions are more important than differences in observable characteristics between sexual preference groups in explaining the income gaps homosexuals face. That is because although equalizing skills distributions would be unable to reduce the differences in sorting into college and white collar occupations, it would increase college educated homosexuals' income proportionally more.

Also, this Chapter shows Norwegian white collar occupations pay significantly more than blue collar ones. Hence, the determinants of taste for white collar occupations are very important in determining annual income. My findings indicate that—leaving occupational choices unaltered—observable characteristics are key determinants of the occupational choices of homosexuals. Therefore, equalizing these traits to those of heterosexuals would increase the take-up of white collar occupations by members of sexual minorities. The take-up increase is so large that under this scenario homosexuals would be more likely than heterosexuals to work in white collar occupations. The new sorting would cause income gaps to disappear and, in fact, it would yield a labor market where homosexuals have an income advantage over heterosexuals.

This is the first work to incorporate endogenous decisions and unobserved skills in the analysis of income discrimination agains homosexuals. My results show that all of these elements are of great importance and that closing gaps in traits and tastes for occupations would significantly reduce the income gaps non-heterosexuals currently face. This topic, of great importance as barriers against sexual minorities continue to fall in other aspects of life, need further analysis. Although the YiN data set used in this Chapter is a great source of information for these topics, the research community would benefit from data sources with a greater sample size. More observations are needed to incorporate education and occupational choices simultaneously in one model, and in that way be able to jointly analyze the two effects above exposed.

BIBLIOGRAPHY

Allegretto, S. and Arthur, M. (2001). An empirical analysis of homosexual/heterosexual male earnings differentials: unmarried and unequal? *Industrial and labor relations review*, pages 631–646.

Angrist, J. (2014). The perils of peer effects. *NBER Working Paper*, (19774).

Antecol, H., Jong, A., and Steinberger, M. (2007). The sexual orientation wage gap: The role of occupational sorting and human capital, and discrimination. *Indus. & Lab. Rel. Rev.*, 61:518.

APA (2006). The american psychological association dictionary. *American Psychological Association Reference Books*.

Badgett, M. (1995). The wage effects of sexual orientation discrimination. *Industrial and labor relations review*, pages 726–739.

Bartholomew, D., Knott, M., and Moustaki, I. (2011). Latent variable models and factor analysis: A unified approach. *Wiley Series in Probability and Statistics. West Sussex, United Kingdom*.

Becker, G. (1957). The economics of discrimination. University of Chicago Press: Chicago.

Becker, G. and Chiswick, B. (1966). Education and the distribution of earnings. *The American Economic Review*, 56(1/2):358–369.

Berg, N. and Lien, D. (2002). Measuring the effect of sexual orientation on income: evidence of discrimination? *Contemporary economic policy*, 20(4):394–414.

Björkqvist, K., Ekman, K., and Lagerspetz, K. (1982). Bullies and victims: Their ego picture, ideal ego picture and normative ego picture. *Scandinavian Journal of Psychology*, 23(1):307–313.

Black, D., Makar, H., Sanders, S., and Taylor, L. (2003). The earnings effects of sexual orientation. *Industrial and labor relations review*, pages 449–469.

Borghans, L., Duckworth, A. L., Heckman, J. J., and Weel, B. T. (2008). The economics and psychology of personality traits. *Journal of Human Resources*, 43(4):972–1059.

Boulton, M. and Underwood, K. (1993). Bully/victim problems among middle school children. *European education*, 25(3):18–37.

Brown, S. and Taylor, K. (2008). Bullying, education and earnings: evidence from the national child development study. *Economics of Education Review*, 27(4):387–401.

Bukowski, W. M., Hoza, B., and Boivin, M. (1995). Friendship protects "at risk" children from victimization by peers. *The role of friendship in children's developmental risk and resilience: A developmental pathology perspective*.

Cabrera, N., Shannon, J., and Tamis-LeMonda, C. (2007). Fathers' influence on their children's cognitive and emotional development: From toddlers to pre-k. *Applied Development Science*, 11(4):208–213.

Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica: Journal of the Econometric Society*, 5:1127–1160.

Carneiro, P., Hansen, K., and Heckman, J. (2003). Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. *International Economic Review*, 44(2):361–422.

Carpenter, C. (2007). Revisiting the income penalty for behaviorally gay men: Evidence from nhanes iii. *Labour economics*, 14(1):25–34.

Carpenter, C. (2009). Sexual orientation and outcomes in college. *Economics of Education Review*, 28(6):693–703.

Cawley, J., Heckman, J., and Vytlacil, E. (2001). Three observations on wages and measured cognitive ability. *Labour Economics*, 8(4):419–442.

Clain, S. and Leppel, K. (2001). An investigation into sexual orientation discrimination as an explanation for wage differences. *Applied economics*, 33(1):37–47.

Cunha, F. and Heckman, J. (2007). The technology of skill formation. *IZA Discussion Paper Series*, 2550.

Cunha, F., Heckman, J., Lochner, L., and Masterov, D. (2006). Interpreting the evidence on life cycle skill formation. *Handbook of the Economics of Education*, 1:697–812.

Cunha, F., Heckman, J., and Schennach, S. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931.

Cunha, F. and Heckman, J. J. (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources*, 43(4):738–782.

Dake, J., Price, J., and Telljohann, S. (2003). The nature and extent of bullying at school. *Journal of school health*, 73(5):173–180.

Doyle, O., Harmon, C., Heckman, J., and Tremblay, R. (2009). Investing in early human development: timing and economic efficiency. *Economics & Human Biology*, 7(1):1–6.

Duckworth, A. and Seligman, M. (2005). Self-discipline outdoes iq in predicting academic performance of adolescents. *Psychological Science*, 16(12):939–944.

Eriksen, T., Nielsen, H., and Simonsen, M. (2012). The effects of bullying in elementary school. *Economics Working Papers*, 16.

Espinoza, R., Sarzosa, M., and Urzua, S. (2014). The causal effect of skills: An international study. *University of Maryland*, pages 1–53.

Faris, R. and Felmlee, D. (2011). Status struggles network centrality and gender segregation in same-and cross-gender aggression. *American Sociological Review*, 76(1):48–73.

Flouri, E. and Buchanan, A. (2002). Life satisfaction in teenage boys: The moderating role of father involvement and bullying. *Aggr. Behav.*, 28(2):126–133.

Fredriksson, P., Ockert, B., and Oosterbeek, H. (2013). Long-term effects of class size. *Quarterly Journal of Economics*.

Hansen, K., Heckman, J., and Mullen, K. (2004). The effect of schooling and ability on achievement test scores. *Journal of Econometrics*, 121(1):39–98.

Hart, B. and Risley, T. (1995). *Meaningful Differences in the Everyday Experience of Young American Children*.

Heckman, J., Humphries, J., and Mader, N. (2011a). The ged. *Handbook of Economics of Education*, 3.

Heckman, J., Humphries, J. E., Urzua, S., and Veramendi, G. (2011b). The effects of educational choices on labor market, health, and social outcomes. *Human Capital and Economic Opportunity: A Global Working Group*, (2011-002):1–63.

Heckman, J. and Masterov, D. (2007). The productivity argument for investing in young children. *Applied Economic Perspectives and Policy*, 29(3):446–493.

Heckman, J. and Rubinstein, Y. (2001). The importance of noncognitive skills: Lessons from the ged testing program. *The American Economic Review*, 91(2):145–149.

Heckman, J., Stixrud, J., and Urzua, S. (2006a). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24(3):411–482.

Heckman, J., Stixrud, J., and Urzua, S. (2006b). Web supplement to "the effects of cognitive and noncognitive abilities on labor market outcomes and social behavior". *jenni.uchicago.edu/noncog*, pages 1–64.

Heckman, J., Urzua, S., and Vytlacil, E. (2006c). Understanding instrumental variables in models with essential heterogeneity. *Review of Economics and Statistics*.

Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., and Yavitz, A. (2010). The rate of return to the highscope perry preschool program. *Journal of Public Economics*, 94(1-2):114–128.

Hodges, E., Malone, M., and Perry, D. (1997). Individual risk and social risk as interacting determinants of victimization in the peer group. *Developmental psychology*, 33(6):1032.

Hodges, E. and Perry, D. (1996). Victims of peer abuse: An overview. *Relaiming Chldren and youth*, Spring:23–28.

Jackson, C. (2013). Non-cognitive ability, test scores, and teacher quality: Evidence from 9th grade teachers in north carolina. *works.bepress.com*.

Judd, K. L. (1998). *Numerical Methods in Economics*.

Kang, C. (2007). Classroom peer effects and academic achievement: Quasi-randomization evidence from south korea. *Journal of Urban Economics*, 61(3):458–495.

Kiernan, K. E. and Huerta, M. C. (2008). Economic deprivation, maternal depression, parenting and children's cognitive and emotional development in early childhood1. *The British Journal of Sociology*, 59(4):783–806.

Kim, Y. S. (2005). School bullying and suicidal risk in korean middle school students. *PEDIATRICS*, 115(2):357–363.

Knudsen, E., Heckman, J., Cameron, J., and Shonkoff, J. (2006). Economic, neurobiological, and behavioral perspectives on building america's future workforce. *Proceedings of the National Academy of Sciences*, 103(27):10155–10162.

Kotlarski, I. (1967). On characterizing the gamma and the normal distribution. *Pacific Journal of Mathematics*, 20(1):69–76.

Ladd, G. and Ladd, B. (1998). Parenting behaviors and parent–child relationships: Correlates of peer victimization in kindergarten? *Developmental psychology*, 34(6):1450.

Laumann, E., Gagnon, J., Michael, R., and Michaels, S. (1994). The social organization of sexuality: Sexual practices in the united states. Chicago: University of Chicago Press.

Le, A., Miller, P., Heath, A., and Martin, N. (2005). Early childhood behaviours, schooling and labour market outcomes: Estimates from a sample of twins. *Economics of Education Review*, 24(1):1–17.

Lowenstein, L. F. (1978). The bullied and non-bullied child. *Bulletin of the British Psychological Society*, (31):316–318.

Madsen, K. C. (1996). Differing perceptions of bullying and their practical implications. *Education and Child Psychology*, 13(2):14–22.

Manski, C. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies*, 60(3):531–542.

Manski, C. (2000). Economic analysis of social interactions. *Journal of Economic Perspectives*, 14(3):115–136.

McEwen, B. and Seeman, T. (2006). Protective and damaging effects of mediators of stress: elaborating and testing the concepts of allostasis and allostatic load. *Annals of the New York Academy of Sciences*, 896(1):30–47.

Mincer, J. (1958). Investment in human capital and personal income distribution. *Journal of Political Economy*, 66(4):281–302.

Mincer, J. and Polachek, S. (1974). *Family investments in human capital: Earnings of women*.

Murnane, R., Willett, J., and Levy, F. (1995). The growing importance of cognitive skills in wage determination. *The Review of Economics and Statistics*, 77(2):251–266.

Naylor, P., Cowie, H., Cossin, F., Bettencourt, R., and Lemme, F. (2010). Teachers' and pupils' definitions of bullying. *British Journal of Educational Psychology*, 76(3):553–576.

Neal, D. A. and Johnson, W. R. (1996). The role of premarket factors in black-white wage differences. *Journal of Political Economy*, 104(5):869–895.

OECD (2014). Skills for social progress. pages 1–103.

of Education, U. D. (2011). Analysis of state bullying laws and policies. pages 1–202.

Olweus, D. (1997). Bully/victim problems in school: Facts and intervention. *European Journal of Psychology of Education*, 12(4):495–510.

Ouellet-Morin, I., Odgers, C. L., Danese, A., Bowes, L., Shakoor, S., Papadopoulos, A. S., Caspi, A., Moffitt, T. E., and Arseneault, L. (2011). Blunted cortisol responses to stress signal social and behavioral problems among maltreated/bullied 12-year-old children. *BPS*, 70(11):1016–1023.

Payton, J., Weissberg, R., Durlak, J., Dymnicki, A., Taylor, R., Schellinger, K., and Pachan, M. (2008). The positive impact of social and emotional learning for kindergarten to eighth-grade students: Findings from three scientific reviews. technical report. *Collaborative for Academic, Social, and Emotional Learning (NJ1).*

Perry, D., Kusel, S., and Perry, L. (1988). Victims of peer aggression. *Developmental psychology*, 24(6):807.

Pollak, R. (1976). Interdependent preferences. *The American Economic Review*, pages 309–320.

Roy, A. (1951). Some thoughts on the distribution of earnings. *Oxford economic papers*, 3(2):135–146.

Salmivalli, C., Lagerspetz, K., Björkqvist, K., Österman, K., and Kaukiainen, A. (1996). Bullying as a group process: Participant roles and their relations to social status within the group. *Aggr. Behav.*, 22(1):1–15.

Sarzosa, M. and Urzua, S. (2012). Implemeting factor models in stata: the heterofactor command. *MIMEO. University of Maryland*, pages 1–6.

Schelling, T. (1971). Dynamic models of segregation. *Journal of mathematical sociology*, 1(2):143–186.

Schuster, B. (1999). Outsiders at school: The prevalence of bullying and its relation with social status. *Group Processes & Intergroup Relations*, 2(2):175–190.

Skinner, E. and Belmont, M. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of educational psychology*, 85(4):571.

Smith, P. and Brain, P. (2000). Bullying in schools: Lessons from two decades of research. *Aggressive behavior*, 26(1):1–9.

Smith, P., Cowie, H., Olafsson, R., and Liefooghe, A. (2002). Definitions of bullying: A comparison of terms used, and age and gender differences, in a fourteen–country international comparison. *Child development*, 73(4):1119–1133.

Smith, P. and Levan, S. (1995). Perceptions and experiences of bullying in younger pupils. *British Journal of Educational Psychology*, 65(4):489–500.

Smith, P., Madsen, K., and Moody, J. (1999). What causes the age decline in reports of being bullied at school? towards a developmental analysis of risks of being bullied. *Educational Research*, 41(3):267–285.

Smith, P., Talamelli, L., Cowie, H., Naylor, P., and Chauhan, P. (2004). Profiles of non-victims, escaped victims, continuing victims and new victims of school bullying. *British Journal of Educational Psychology*, 74(4):565–581.

Strand, N. P. and von Soest, T. (2008). Young in norway - longitudinal. documentation of design, variables and scales. NOVA.

Tough, P. (2012). How children succeed: Grit, curiosity, and the hidden power of character.

Urzua, S. (2008). Racial labor market gaps. *Journal of Human Resources*, 43(4):919.

Vossekuil, B., Fein, R., Reddy, M., Borum, R., and Modzeleski, W. (2002). The final report and findings of the safe school initiative. *US Secret Service and Department of Education, Washington, DC.*

Wichstrøm, L., von Soest, T., and Kvalem, I. (2013). Predictors of growth and decline in leisure time physical activity from adolescence to adulthood. *Health psychology.*

Willis, R. and Rosen, S. (1979). Education and self-selection. *The Journal of Political Economy*, 87(5):S7–S36.

Wolke, D., Woods, S., Stanford, K., and Schulz, H. (2001). Bullying and victimization of primary school children in england and germany: prevalence and school factors. *British Journal of Psychology*, 92(4):673–696.

APPENDICES

G *Descriptive Statistics of Outcomes*

Tab. G.1: Descriptive Statistics of Outcome and Treatment Variables

Outcome	Mean	Std. Dev.	Min	Max
At $t = 2$				
Bullied	0.1119	0.3153	0	1
Bully	0.0777	0.2677	0	1
Cyberbully	0.2827	0.4503	0	1
Criminal	0.1249	0.3307	0	1
Cybercriminal	0.1276	0.3337	0	1
At $t = 5$				
Depression	15.416	4.524	6	30
Drinking	0.7213	0.4483	0	1
Smoking	0.1907	0.3929	0	1
Life Satisfaction	0.4772	0.4995	0	1
Sick	0.0735	0.2611	0	1
Mental Issues	0.0315	0.1749	0	1
Stressed Image	10.644	3.6013	4	20
Stressed Total	45.471	11.354	17	85

Tab. G.2: Descriptive Statistics of Outcome Variables

All

Variable	Obs	Mean	Std. Dev.	Min	Max	Variable	Obs	Mean	Std. Dev.	Min	Max
Depression	2833	15.4168	4.524	6	30	Volunteer	3449	0.1032	0.3042	0	1
Drinking	3449	0.7213	0.4483	0	1	Beh Problem	3449	0.1249	0.3307	0	1
Smoking	3449	0.1907	0.3929	0	1	Bullied	3188	0.1119	0.3153	0	1
Life Satisfa	3449	0.4772	0.4995	0	1						

Males

Variable	Obs	Mean	Std. Dev.	Min	Max	Variable	Obs	Mean	Std. Dev.	Min	Max
Depression	1348	14.566	4.4594	6	30	Volunteer	1725	0.0794	0.2704	0	1
Drinking	1725	0.6904	0.4624	0	1	Beh Problem	1725	0.1426	0.3497	0	1
Smoking	1725	0.2846	0.4513	0	1	Bullied	1594	0.1373	0.3443	0	1
Life Satisfa	1725	0.5002	0.5001	0	1						

Females

Variable	Obs	Mean	Std. Dev.	Min	Max	Variable	Obs	Mean	Std. Dev.	Min	Max
Depression	1485	16.1892	4.4444	6	29	Volunteer	1724	0.127	0.3331	0	1
Drinking	1724	0.7523	0.4317	0	1	Beh Problem	1724	0.1073	0.3095	0	1
Smoking	1724	0.0968	0.2958	0	1	Bullied	1594	0.0865	0.2812	0	1
Life Satisfa	1724	0.4541	0.498	0	1						

Attending College

Variable	Obs	Mean	Std. Dev.	Min	Max	Variable	Obs	Mean	Std. Dev.	Min	Max
Depression	1954	15.3756	4.4583	6	30	Life Satisfa	1954	0.6084	0.4882	0	1
Drinking	1954	0.9002	0.2998	0	1	Volunteer	1954	0.1606	0.3673	0	1
Smoking	1954	0.196	0.397	0	1						

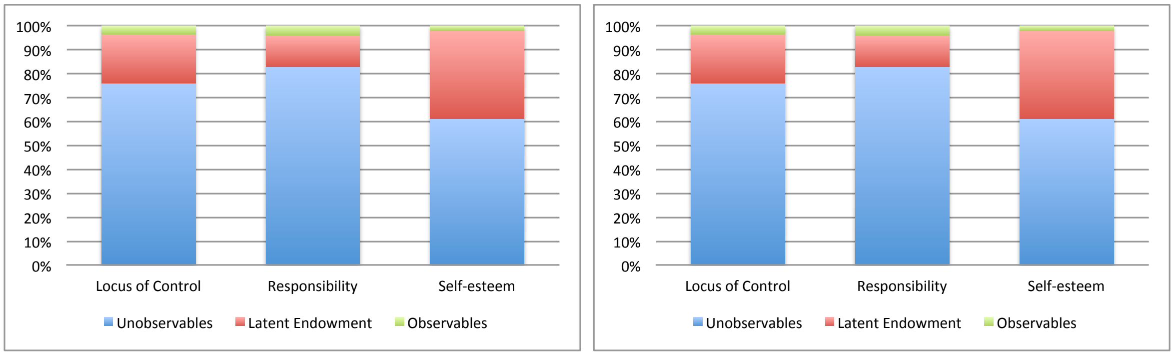
Not Attending College

Variable	Obs	Mean	Std. Dev.	Min	Max	Variable	Obs	Mean	Std. Dev.	Min	Max
Depression	879	15.5085	4.6679	6	30	Life Satisfa	1495	0.3056	0.4608	0	1
Drinking	1495	0.4876	0.5	0	1	Volunteer	1495	0.028	0.1652	0	1
Smoking	1495	0.1839	0.3875	0	1						

H Variance Decomposition of Test Scores, Orthogonal Skills Model

Fig. H.1: Variance Decomposition of Test Scores

(a) Non-Cognitive Test Scores *(b)* Cognitive Test Scores



I Information Used to create Non-cognitive score

In the case of locus of control, we added three questions:

1. I have confidence in my own decision
2. I believe that I can deal with my problems by myself
3. I am taking full responsibility of my own life

To create the self-esteem index we added:

1. I think that I have a good character
2. I think that I am a competent person
3. I think that I am a worthy person
4. Sometimes I think that I am a worthless person (the negative of)
5. Sometimes I think that I am a bad person (the negative of)
6. I generally feel that I am a failure in life (the negative of)
7. If I do something wrong, people around me will blame me much (the negative of)
8. If I do something wrong, I will be put to shame by people around me (the negative of)

Finally we created the irresponsibility index by adding:

1. I jump into exciting things even if I have to take an examination tomorrow
2. I abandon a task once it becomes hard and laborious to do
3. I am apt to enjoy risky activities

J Attrition Analysis

In this Appendix, I present some estimations regarding the observations lost due to attrition in the data set used in Chapters 1 and 2. First, Table J.1 shows that attrition in the first three waves is relatively low compared with similar surveys.

Tab. J.1: Attrition by Wave

	Attrit.	Attr. if Bull in $t = 1$	
Wave		No	Yes
1	.	.	.
2	7.5%	7.7%	7%
3	9.4%	9.1%	10.4%
4	9.5%	9.1%	11%
5	13.9%	13.7%	14.8%

In addition, Tables J.2 and J.3 show that there are few differences between those who leave the sample and those who stay. The only observable characteristics in which the attrited and the non-attrited subsamples differ are income, the proportion of fathers with graduate school and two of the cognitive tests. These differences are significant at the 90% confidence level. It is important to note that there are no statistical differences between the subsamples according to bullying perpetration or victimization. In Table J.3, I incorporate the unobservables (i.e., cognitive and non-cognitive skills) in the analysis. It shows that, consistent with the findings in Table J.2, the kids that leave the sample are low cognitive skilled wealthy kids with highly educated parents.

Tab. J.2: Difference in Observables at $t = 1$ of Attrited and Non-Attrited Observations

Variable	Mean Att	Mean Stay	Diff.
Age (months)	8.6346	8.9626	-.328
Male	.5019	.5	.0019
Older Siblings	.4559	.5452	-.0893*
Young Siblings	.6398	.6341	.0058
lnInc_pc	4.5632	4.3275	.2356*
Urban	.8659	.8676	-.0017
Lives: Both Parents	.9195	.9294	-.0099
Lives: Only Mother	.0383	.0332	.0051
Father Edu: 2yColl	.0728	.0678	.005
Father Edu: 4yColl	.295	.2974	-.0023
Father Edu: GS	.1341	.063	.0711*
Locus of Control	.0631	-.0052	.0682
Irresponsibility	-.0827	.0068	-.0895
Self-Esteem	.0068	-.0006	.0074
Language & SocStd	-.0907	.0074	-.0981
Math & Science	-.1457	.0119	-.1576*
Yearly Test	-.108	.009	-.117*
Bullied	.2107	.2262	-.0154
Bully	.2759	.2437	.0321

*** p<0.01, ** p<0.05, * p<0.1

Tab. J.3: Probability of Staying from $t = 1$ to $t = 2$

	Raw Coefficients	
Stay in Wave 2	Coeff.	StdErr.
Age (months)	0.0092	(0.010)
Male	-0.0004	(0.072)
Older Siblings	0.0531	(0.070)
Young Siblings	-0.0287	(0.070)
lnInc_pc	-0.3089***	(0.068)
Urban	0.1250	(0.106)
Lives: Both Parents	0.1375	(0.209)
Lives: Only Mother	-0.1876	(0.273)
Father Edu: 2yColl	-0.0036	(0.146)
Father Edu: 4yColl	-0.0727	(0.085)
Father Edu: GS	-0.4410***	(0.126)
Non-Cognitive	-0.2479	(0.321)
Cognitive	0.1497*	(0.078)
Constant	2.5977***	(0.365)
Observations	3,097	

*** p<0.01, ** p<0.05, * p<0.1

K Proofs

K.1 Proof Theorem 1

Organize the adjunct measurement system (3.3) such that the subset of measures affected only by θ^A remain on the top L_A rows and the rest of the measures remain in the bottom $L_{A,B} = L - L_A$ rows. That way, we can partition the measurement system in two blocks

$$\begin{bmatrix} \mathbf{T}^A \\ \\ \mathbf{T}^{A,B} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_T\beta^T + \alpha^{(A)A}\theta^A + \mathbf{e}^T \\ \\ \mathbf{X}_T\beta^T + \alpha^{(A,B)A}\theta^A + \alpha^{(A,B)B}\theta^B + \mathbf{e}^T \end{bmatrix}$$

Assuming that the latent factor A has non-degenerate distributions

$$COV\left(T_h^A, T_k^A\right) = \alpha_h^{(A)A}\alpha_k^{(A)A}\sigma^2_{\theta^A} \qquad \text{for } h, k = 1, \ldots, L_A; \quad h \neq k$$

and

$$COV\left(T_h^A, T_l^A\right) = \alpha_h^{(A)A}\alpha_l^{(A)A}\sigma^2_{\theta^A} \qquad \text{for } h, l = 1, \ldots, L_A; \quad h \neq l$$

Hence,

$$\frac{COV\left(T_h^A, T_k^A\right)}{COV\left(T_h^A, T_l^A\right)} = \frac{\alpha_k^{(A)A}}{\alpha_l^{(A)A}}$$

Therefore, $L_A - 1$ factor loadings are identified up to one normalization. Also, assuming with out loss of generality that the normalized loading is that of equation l we have that

$$\sigma^2_{\theta^A} = \frac{COV\left(T_h^A, T_l^A\right) COV\left(T_k^A, T_l^A\right)}{COV\left(T_h^A, T_k^A\right)}$$

Now turning to the second block of the measurement system and assuming that latent factor B has a non-degenerate distribution, I have that

$$COV\left(T_m^B, T_n^B\right) = \alpha_m^{(A,B)A}\alpha_n^{(A,B)A}\sigma_{\theta^A}^2 + \alpha_m^{(A,B)B}\alpha_n^{(A,B)B}\sigma_{\theta^B}^2 + \left(\alpha_m^{(A,B)A}\alpha_n^{(A,B)B} + \alpha_m^{(A,B)B}\alpha_n^{(A,B)A}\right)\sigma_{\theta^A,\theta^B} \quad \text{(K.1)}$$

for $m, n = 1, \ldots, L_B$ and $m \neq n$, and

$$COV\left(T_m^B, T_l^A\right) = \alpha_m^{(A,B)A}\alpha_l^{(A)A}\sigma_{\theta^A}^2 + \alpha_m^{(A,B)A}\alpha_l^{(A)A}\sigma_{\theta^A,\theta^B} \quad \text{(K.2)}$$

for $m = 1, \ldots, L_B$ and $l = 1, \ldots, L_A$. It is easy to see that this second block of the measurement system is not identified as it has $2L_B + 2$ unknowns, while it has only $2L_B$ pieces of relevant information. That is, I have two loadings per measure plus $\sigma_{\theta^B}^2$ and $\sigma_{\theta^A,\theta^B}$ to identify, and I have two covariances per each L_B tests: one with another second block test and another one with one the L_A test (preferably, the one that has the normalized loading).

Now, let me reduce the number of unknowns by normalizing one of the second block loadings (i.e., $\alpha_o^{(A,B)B} = 1$ for $o = \{1, .., L_B\}$) and assuming one of the second block measures is only affected by the second factor (i.e., $\alpha_o^{(A,B)A} = 0$), then

$$COV\left(T_m^B, T_o^B\right) = \alpha_m^{(A,B)B}\sigma_{\theta^B}^2 + \alpha_m^{(A,B)A}\sigma_{\theta^A,\theta^B} \quad \text{(K.3)}$$

Using the normalization of $\alpha_l^{(A)A} = 1$ I reported above, (K.2) becomes

$$COV\left(T_m^B, T_l^A\right) = \alpha_m^{(A,B)A}\sigma_{\theta^A}^2 + \alpha_m^{(A,B)B}\sigma_{\theta^A,\theta^B} \quad \text{(K.4)}$$

Furthermore, $COV\left(T_l^A, T_o^B\right) = \sigma_{\theta^A,\theta^B}$. Then, using (K.3) and (K.4), I can identify $\alpha_m^{(A,B)A}$ and $\alpha_m^{(A,B)B}$ as a function of measurement covariances and $\sigma^2_{\theta^B}$, which I can later identify using equation (K.1).

K.2 Proof Theorem 2

To simplify exposition let us reduce the number of parameters to two, and rename ρ as γ_2. Hence, the parameter set is $\mathbf{\Gamma} = \{\gamma_1, \gamma_2\}$. Assuming separability of the error term function to estimate is

$$\theta_{t+1} = h\left(\theta_t, \mathbf{\Gamma}\right) = \left(\gamma_1 \theta_{t,A}^{\gamma_2} + \left(1 - \gamma_1\right) \theta_{t,B}^{\gamma_2}\right)^{1/\gamma_2} + \varepsilon$$

I linearize the CES function using a Taylor approximation around $\mathbf{\Gamma}^0$ in order to rely on the results of the linearized regression model described in Green (2000). That is, if $\theta_k^0 = \partial h\left(\theta_t, \mathbf{\Gamma}\right)/\partial \gamma_k^0$

$$h\left(\theta_t, \mathbf{\Gamma}\right) \backsimeq h\left(\theta_t, \mathbf{\Gamma}^0\right) - \gamma_1^0\theta_1^0 - \gamma_2^0\theta_2^0 + \gamma_1\theta_1^0 + \gamma_2\theta_2^0$$

or if I stack the variables into matrices

$$\theta_{t+1} \backsimeq h\left(\theta_t, \mathbf{\Gamma}^0\right) - \gamma^0\theta^0 + \gamma\theta^0 + \varepsilon$$

Defining ε^0 as the error term that contains the true disturbance ε and the deviation that arises due to the Taylor approximation, and $\theta_{t+1}^0 = \theta_{t+1} - h\left(\theta_t, \mathbf{\Gamma}^0\right) + \gamma^0\theta^0$ then

I have

$$\theta^0_{t+1} = \gamma\theta^0 + \varepsilon^0$$

which can be estimated using least squares. Given that I am particularly interested in γ_2, let me face the estimation procedure using the partitioned regression.

$$\theta^0_{t+1} = \begin{bmatrix} \theta^0_1 & \theta^0_2 \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} + \varepsilon^0$$

From the Frisch-Waugh theorem we now that $\hat{\gamma}_2$ can be obtained by regressing the residuals from the regression between θ^0_{t+1} and θ^0_1 on the residuals from the regression between θ^0_2 and θ^0_1. That is,

$$\hat{\gamma}_2 = \left(\theta^{0\prime}_2 M_{\theta^0_1} \theta^0_2\right)^{-1} \theta^{0\prime}_2 M_{\theta^0_1} \theta^0_{t+1}$$

where $M_{\theta^0_1} = I - \theta^0_1 \left(\theta^{0\prime}_1 \theta^0_1\right)^{-1} \theta^{0\prime}_1$ is the annihilator matrix of the regression of θ^0_{t+1} on θ^0_1. Hence, I can check the bias in $\hat{\gamma}_2$

$$\begin{aligned}
E\left[\hat{\gamma}_2 | \theta^0_1, \theta^0_2\right] &= E\left[\left(\theta^{0\prime}_2 M_{\theta^0_1} \theta^0_2\right)^{-1} \theta^{0\prime}_2 M_{\theta^0_1} \left[\theta^0_1\gamma_1 + \theta^0_2\gamma_2 + \varepsilon^0\right]\right] \\
&= \left(\theta^{0\prime}_2 M_{\theta^0_1} \theta^0_2\right)^{-1} \theta^{0\prime}_2 \left[M_{\theta^0_1} \left[\theta^0_1\gamma_1 + \theta^0_2\gamma_2\right] + M_{\theta^0_1} E\left[\varepsilon^0\right]\right] \\
&= \left(\theta^{0\prime}_2 M_{\theta^0_1} \theta^0_2\right)^{-1} \theta^{0\prime}_2 \left[M_{\theta^0_1} \theta^0_2\gamma_2 + M_{\theta^0_1} E\left[\varepsilon^0\right]\right] \\
&= \gamma_2 + \left(\theta^{0\prime}_2 M_{\theta^0_1} \theta^0_2\right)^{-1} \theta^{0\prime}_2 M_{\theta^0_1} E\left[\varepsilon^0\right]
\end{aligned}$$

Therefore the, bias in $\hat{\gamma}_2$ depends on the mean of ε^0 which in turn depends on the mean of θ_{t+1} trough ε. Hence, the only way $E\left[\hat{\gamma}_2\right] = \gamma_2$ is if $E\left[\theta_{t+1}\right] = 0$.

L Information used to create the measures of investment in non-cognitive skills

In the creation of the non-cognitive investment measures I used several variables and combined them in three indexes, namely parental abuse, parental control and parental harmony.

The parental abuse index is an aggregation of the following variables:

- I frequently see my parents verbally abuse each other
- I frequently see one of my parents beat the other one
- I am often verbally abused by parents
- I am often severely beaten by parents

The parental control index is created by aggregating:

- When I go out, my parents usually know where I am
- When I go out, my parents usually know whom I am with
- When I go out, my parents usually know what I do
- When I go out, my parents usually know when I return

Finally, the parental harmony index is created using the following variables:

- My parents and I try to spend much time together
- My parents always treat me with love and affection
- My parents and I understand each other well
- My parents and I candidly talk about everything
- I frequently talk about my thoughts and what I experience away from home with my parents
- My parents and I have frequent conversations

Tab. L.1: Descriptive Statistics of Investment Indexes

	Parental Abuse		Parental Control		Parental Harmony	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
All	7.1641	3.157	0.002	0.997	0	0.999
Males	7.1627	3.094	-0.118	0.983	-0.086	0.963
Females	7.1654	3.219	0.124	0.997	0.085	1.025
Attending College*	6.8914	3.008	0.038	1.012	0.036	0.969
Not Attending College*	7.5211	3.309	-0.043	0.098	-0.048	1.034

* Sample limited to wave 6

M Outcomes Analysis at age 16

In this Appendix I present some results that help understand the impacts found in the Chapter 2 using understandable metrics. I estimate the following specification:

$$Y = \mathbf{X}_Y \beta^Y + \alpha^{Y,NC} \theta^{NC} + \alpha^{Y,C} \theta^C + e^Y$$

Its purpose is to capture the effect of skills on more tangible outcomes, and in that way have a better picture about how the skills lost to bullying hurt the development of successful lives. See Chapter 1 for a detailed explanation on how the outcome measures were constructed.

Tab. M.1: Effect of unobserved heterogeneity at age 16

	(1)	(2)	(3)	(4)	(5)
			Probit		
VARIABLES	College	LifeSatisf	Healthy	Smoke	Drink
Age (months)	0.0040	0.0139	0.0079	-0.0002	-0.0125*
	(0.008)	(0.011)	(0.007)	(0.010)	(0.007)
Male	-0.3658***	0.3862***	0.0930*	0.4307***	-0.0347
	(0.057)	(0.093)	(0.054)	(0.074)	(0.052)
Older Siblings	-0.0005	0.1973**	0.0333	-0.0918	-0.0197
	(0.054)	(0.082)	(0.052)	(0.068)	(0.051)
Young Siblings	0.0405	0.1529*	-0.0233	-0.1841**	-0.0967*
	(0.055)	(0.082)	(0.053)	(0.073)	(0.051)
lnInc_pc	0.0811	0.3043***	0.1302**	0.0366	-0.1029**
	(0.056)	(0.088)	(0.053)	(0.067)	(0.052)
Urban	-0.1443*	0.0536	-0.1025	0.1205	-0.0566
	(0.086)	(0.118)	(0.079)	(0.108)	(0.077)
Lives: Both Parents	0.3620***	0.0954	0.3392***	-0.3425**	-0.3686***
	(0.132)	(0.190)	(0.130)	(0.147)	(0.120)
Lives: Only Mother	0.2993	0.3979	0.2900	0.1326	0.0274
	(0.185)	(0.269)	(0.180)	(0.196)	(0.168)
Father Edu: 2yColl	0.1782	0.2736*	-0.2132**	-0.4097**	-0.1278
	(0.119)	(0.161)	(0.104)	(0.164)	(0.103)
Father Edu: 4yColl	-0.0196	0.2157**	-0.0199	-0.1406*	-0.0358
	(0.068)	(0.099)	(0.064)	(0.084)	(0.062)
Father Edu: GS	-0.1255	0.5521***	0.0177	-0.4167**	-0.1610
	(0.125)	(0.200)	(0.120)	(0.187)	(0.122)
ParentWantsColl	0.6257***	0.1989	0.0858	-0.2264**	-0.0391
	(0.090)	(0.136)	(0.090)	(0.107)	(0.087)
ParentWantsGS	0.5497***	0.2426	0.0218	-0.2332	-0.0226
	(0.115)	(0.172)	(0.112)	(0.143)	(0.109)
Non-Cogs	-0.0726	3.5356***	1.0919***	-0.6746***	-0.5552***
	(0.169)	(0.676)	(0.190)	(0.231)	(0.163)
Cognitive	0.0893**	-0.2060***	-0.1937***	-0.1479***	-0.0570
	(0.041)	(0.077)	(0.043)	(0.045)	(0.037)
Constant	-0.4744	-2.1872***	-1.0961***	-1.0815***	0.7336***
	(0.296)	(0.502)	(0.291)	(0.357)	(0.276)
Observations	2,345	2,685	2,685	2,685	2,685

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

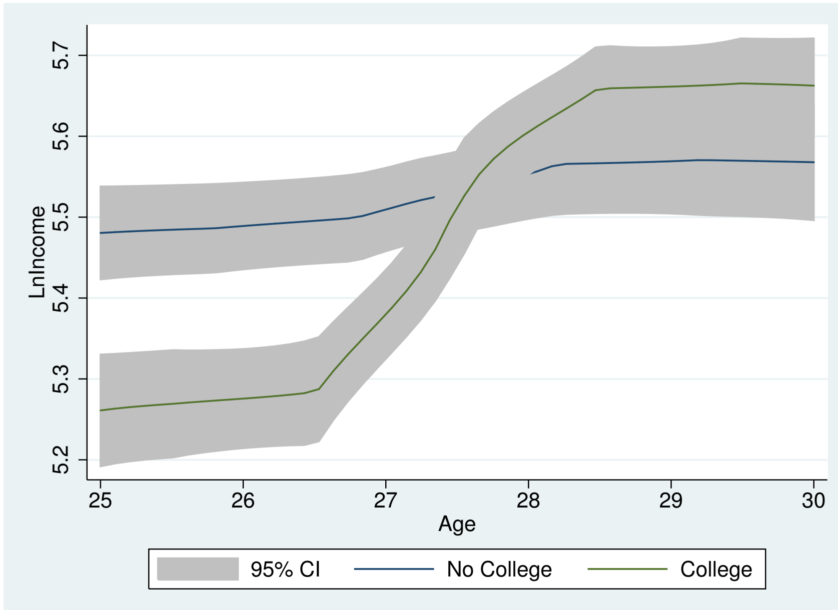
Tab. M.2: Effect of unobserved heterogeneity at age 16

	(6)	(7)	(8)	(9)	(10)
			LS		
VARIABLES	Depression	StressImage	StressFiends	StressSchool	StressTotal
Age (months)	-0.0009	0.0045	0.0054	-0.0023	0.0023
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Male	-0.2526***	-0.3412***	0.0875**	-0.2591***	-0.1685***
	(0.037)	(0.037)	(0.039)	(0.037)	(0.037)
Older Siblings	-0.0094	0.0224	0.0463	0.0641*	0.0408
	(0.036)	(0.036)	(0.038)	(0.036)	(0.036)
Young Siblings	-0.0013	-0.0172	0.0492	0.0726**	0.0284
	(0.036)	(0.037)	(0.038)	(0.037)	(0.037)
lnInc_pc	0.0073	-0.0139	0.0330	0.0574	0.0319
	(0.037)	(0.037)	(0.038)	(0.037)	(0.037)
Urban	-0.0354	-0.0007	-0.1488***	0.1336**	0.0273
	(0.054)	(0.055)	(0.057)	(0.055)	(0.055)
Lives: Both Parents	-0.0692	-0.1907**	-0.1186	0.1786**	-0.0420
	(0.088)	(0.088)	(0.091)	(0.088)	(0.088)
Lives: Only Mother	-0.0950	-0.3407***	-0.2475*	-0.0591	-0.3171***
	(0.121)	(0.123)	(0.127)	(0.122)	(0.123)
Father Edu: 2yColl	0.0093	0.0032	0.0091	0.1268*	0.0910
	(0.072)	(0.072)	(0.075)	(0.072)	(0.073)
Father Edu: 4yColl	0.0272	-0.0990**	-0.0505	0.0910**	0.0013
	(0.044)	(0.044)	(0.046)	(0.044)	(0.044)
Father Edu: GS	-0.0339	-0.2672***	-0.1435	0.0037	-0.1677**
	(0.084)	(0.085)	(0.087)	(0.084)	(0.085)
ParentWantsColl	0.0237	0.0287	-0.0082	0.4518***	0.1969***
	(0.061)	(0.062)	(0.064)	(0.062)	(0.062)
ParentWantsGS	0.1340*	0.0805	0.0956	0.5739***	0.3283***
	(0.077)	(0.078)	(0.081)	(0.078)	(0.078)
Non-Cogs	-1.9588***	-1.5315***	-1.2128***	-1.3627***	-1.7103***
	(0.111)	(0.110)	(0.114)	(0.104)	(0.106)
Cognitive	0.2313***	0.1870***	0.1199***	0.1079***	0.1391***
	(0.026)	(0.027)	(0.028)	(0.026)	(0.026)
Constant	0.1578	0.3884*	-0.0374	-0.9150***	-0.2890
	(0.196)	(0.199)	(0.205)	(0.197)	(0.198)
Observations	2,685	2,676	2,678	2,678	2,654

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

N Additional Tables and Graphs of Chapter 3

Fig. N.1: ln Annual Income by Age and Schooling Level



Note: Local polynomial regressions of (ln) annual income for employed population and age.

Tab. N.1: Difference in Observable Characteristics Between Heterosexuals and Homosexuals

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
			Father's Education			Father's SES			When Child Lived With:		
	Age	Gender	Univers	3YrColl	Vocat.	High	Skilled	Sibling	OnePar	RemPar	NoPar
SI	-0.050	0.281***	0.042**	0.015	-0.010	0.053**	-0.022	-0.032	0.027*	0.024**	0.001
	(0.095)	(0.025)	(0.018)	(0.015)	(0.018)	(0.023)	(0.018)	(0.058)	(0.016)	(0.011)	(0.011)
SP	-0.317***	0.200***	0.014	-0.022	-0.019	-0.021	-0.019	0.078	0.042**	0.023*	0.004
	(0.099)	(0.026)	(0.019)	(0.015)	(0.019)	(0.024)	(0.019)	(0.060)	(0.017)	(0.012)	(0.011)
SX	-0.121	0.071**	0.067***	-0.014	-0.027	0.039	-0.050**	0.031	0.060***	0.027*	0.024*
	(0.121)	(0.032)	(0.023)	(0.019)	(0.023)	(0.029)	(0.023)	(0.073)	(0.021)	(0.014)	(0.014)
SX+	-0.266***	0.165***	0.014	-0.023	-0.029	-0.020	-0.019	0.071	0.057***	0.021*	0.012
	(0.102)	(0.027)	(0.020)	(0.016)	(0.019)	(0.025)	(0.019)	(0.062)	(0.018)	(0.012)	(0.012)

Note: Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. All estimations were drawn from different regressions with the dependent variable being the homosexuality indicator. *Univers* estimates whether the father went to the university. *3YRColl* estimates whether the father went to a three-year college. *Vocat* estimates whether the father undertook vocational training. *OnePat* estimates if the respondent lived with only one of the two parents. *RemPar* estimates if the respondent lived with one father that remarried. *NoPar* estimates whether the respondent lived without her parents during childhood.