The Long-Run Effects of a Bad Start

ISBN: 978 90 5170 7823

This book is no. **647** of the Tinbergen Institute Research Series, established through cooperation between Rozenberg Publishers and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back. (end of this page)

The Long-Run Effects of a Bad Start

De lange termijn effecten van een slechte start

Thesis

to obtain the degree of Doctor from the

Erasmus University Rotterdam

by command of the

rector magnificus

Prof.dr. H.A.P. Pols

and in accordance with the decision of the Doctorate Board.

The public defence shall be held on

Thursday 21 April 2016 at 11:30 hours

by

VIOLETA VALERIEVA MISHEVA

born in Blagoevgrad, Bulgaria

Erasmus University Rotterdam

Erasmus

Doctoral Committee

Promotor:
Prof.dr. H.D. Webbink

Other members:
Prof.dr. A.J. Dur

Prof.dr. M. Lindeboom

Prof.dr. T.G.M. van Ourti

Copromotors:
Dr. A.C. Gielen

Acknowledgement

In the following lines, I would like to thank everyone who helped me along the way of writing and finishing this dissertation.

First, none of this would be possible without my PhD supervisor Prof. Dinand Webbink, who first of all gave me the chance to do my research under his supervision and most of all, always provided me with tireless comments, suggestions and support during the whole process of writing this thesis. He always allowed and encouraged me to pursue my own ideas and guided me towards their refinement. I would also like to thank my co-supervisor Dr. Anne Gielen for agreeing to be a part of the process at a very short notice and for her detailed comments on how to improve the dissertation. I would also like to thank the committee members for reading my dissertation in due time and providing me with comments on possible editions. I am also grateful to Prof. Maarten Lindeboom who, when I was not able to do my research under his supervision, took it as his responsibility to help me find a new supervisory match.

I am grateful to the Tinbergen Institute and Erasmus School of Economics for creating such a high quality academic environment, for the many seminars and interesting guests, and for the constant chance for self-growth and improvement.

I have learned a lot and gained valuable personal and professional experiences from international travels and visits. I am very happy I managed to present my research at high-quality conferences, meet like-minded people and learn so much from them. I am still making a good use of the fancy Ipod I won for my paper during my first ever conference. I am also grateful to many scholars at the Netherlands Institute for the Study of Crime and Law Enforcement not only for providing me with data I used in one of the chapters but also for giving me additional assistance and support during my numerous inquiries, and having their doors open for me to use their facilities in order to try my research ideas.

My first specialty is not economics but journalism. That's why when from VoxEU.org, the website of the Centre for Economic and Policy Research, were looking for PhD students who would work as assistant editors, I was excited to apply. They appreciated my

enthusiasm (and the combination of my degrees) and I have been involved with them for the past couple of years. In that job, I have learned a lot about following strict deadlines, and most of all, I had the chance to read and edit articles written in all areas of economics by some of the leading researchers in the respective fields.

None of the struggle and perseverance required in doing a PhD would be possible without a strong support system. I am very grateful to my friends and colleagues from the Tinbergen Institute and Erasmus University for being such a lovely, and sometimes corky, but never boring group of smart people. In particular, I am thankful to Uyanga for our many *lovely* evenings, dinners, parties, movies and TV series, sport events, and all, and for always being an attentive and compassionate listener who provided me with rational advice when I myself felt lost. The rest of the Erasmus-TI gang: Shawny, Tong, Sait, Ilke – I so enjoyed our dinners in the Pavillion, I will miss them tremendously! I share and appreciate great memories and experiences with many of my other TI colleagues, in particular Eszter, Gosia, Egle, Arturas, Marius, Wooky, Lisette, Steffen, Max, Tomasz, Piotr, Marcin, Lydia and others (whom I will be terrified to realize later that I have forgotten). I also appreciate how my friends out of academia managed to bring me to reality and helped me escape living in a bubble, and I've also met many lovely people in my Dutch classes, during my boxing training, running, and just everyday life. I also appreciate having in my life some lifelong friends who still live in Bulgaria. Their support, especially when I first moved to the Netherlands, has been so valuable to me!

I would also like to thank my parents for their support throughout my whole life, for never putting any pressure and expectations on me, but for always being proud of me and believing in me. I know it was not easy for them to support my choice to move far from home but somehow they managed to accept my ambitions and drives.

Last but not least, I would like to thank Elwin for always believing in me, and for making me believe in myself as well. Thank you for showing interest in my research (and for helping me translate my Dutch summary). I am grateful for the happy moments we have shared over the past few years and for the way you challenged me and helped me

grow as a person. I am looking forward to more of that in the future!

Contents

Chapter 1

Introduction

> "You are the sum total of everything you've ever seen, heard, eaten, smelled, been told, forgot- it's all there. Everything influenced each of us... " Maya Angelou

Having a bad start in life can take different forms with numerous long-term outcomes. For example, being born in a given country at a certain time could be directly related to poor material surroundings, lack of personal freedom and resources, lack of proper education and healthcare, and of an overall opportunity to pursue a better future. Alternatively, negative experiences could occur due to one's unfortunate placement within a society, a community, or a family. Different traumatic experiences can shape an individual's life path by affecting a person's human capital accumulation, career path, choice of peers, criminal behavior, and overall well-being. Is a person indeed an accumulation of all of his experiences, as the above quote suggests, and what can be done to recoup the incurred damages? In the following chapters, we look at the way negative experiences – very often taking place during childhood – affect a person's life.

In particular, we look in detail at problems such as what happens in a person's life so that he or she chooses a criminal career instead of building human capital as a regular

participant in the labor market. Child matreatment is commonly associated with criminal attachment. In a broader context, studies show that if someone has been a victim of a crime he or she is more likely to seek revenge and offend in the future. We explore these problems in Chapters 2 and 3, respectively. One way to reduce criminal involvement is through investment in education. But is the educational system in place effective at instilling skills and knowledge in pupils equally? Or are the relative age placement within a grade and socio-economic background important in educational accumulation? In Chapter 4, we investigate how being the youngest among one's class peers affects the early stages of human capital accumulation. Finally, we look at the way traumatic experiences affect one's well-being. How harmful are different negative experiences and traumatic events in reducing one's quality of life? And after a tragedy, how much time do people need to return to a baseline level of happiness?

All of these questions are important because they are associated with large social costs. Governments around the world pay millions to fight and prevent crime. In the Netherlands and Australia, which are two countries we use in our analyses, crime costs amount to about 4 or 5% of the GDP. Loss of quality of life, though difficult to translate into monetary terms, also bears large costs. Therefore, it is not surprising that coping with different unfavorable events in people's lives has been a central goal of policymakers and academics alike. However, to better direct policy, we need to have a deeper understanding of why certain events happen and how they affect the people who experience them. Very often due to the complexity of a certain problem, policy prescriptions are not built on precise quantitative evidence. The current evidence in social sciences on the issues we discuss is abundant but mostly correlational. Studies establish, for instance, that victims of child abuse more frequently grow up to become criminals, or that offenders more often than non-offenders were victims in the past (Currie & Tekin (2012), Widom (1989), Widom & Maxfield (2001), Widom (1992),Wolfgang & Ferracuti (1967), Ousey *et al.* (2011), Sampson & Lauritsen (1990), Smith & Ecob (2007)). It is also a common finding that relatively younger students perform worse on test scores than their older peers, or that the trau-

matic experiences reduce people's well-being (Bickel *et al.* (1991), Bedard & Dhuey (2006), Fredricksson & Ockert (2013), Muehlenweg & Puhani (2010), Stipek (2002)). However, most studies exploring these issues do not precisely quantify their findings as they cannot isolate the effect of the negative experience from the influence of other factors. There could be many other measured or unmeasurable factors that affect the outcome. The tradition in economic studies in the past few decades has been focused on causal inference. Economists have mastered and perfected different techniques that circumvent the need to use a randomized experiment, using different natural experiments or quasi-experiments when possible, or developing new econometric methods. This is important because many problems cannot be studied in a randomized experimental setting, making causal analysis quite difficult. Following the economic tradition, in the next chapters we use various techniques and approaches to provide a compelling causal evidence of the long-run effects of negative experiences.

Child abuse, as shown in social science research, has many detrimental long-run effects on the victims, such as criminal involvement, depression, and drug dependence, among others. In Chapter 2, we look at the long-run effect of physical and sexual child maltreatment on later problematic outcomes using a sample of twins. Since randomized experiments related to child abuse are not feasible, we rely on twins data as a natural experiment to tackle the issue. Twins are not only genetically similiar but are born in the same family at the same time, thus also share the same family circumstances. We find strong evidence that childhood maltreatment has a large effect on illegal and problematic behavior later in life. The estimated effects imply an increase of illegal or problematic behavior of fifty to one hundred percent relative to the baseline levels of individuals that have not been maltreated. Physical and sexual maltreatment have a large effect on all types of illegal or problematic behavior. Sexual maltreatment by an outsider increases externalizing behavior whereas sexual maltreatment by a family member induces internalizing behavior. A shortcoming of this approach is that though we reduce the endogeneity involving child maltreatment, we do not remove it completely as a randomized experiment or an instru-

ment within such a setting is not feasible. Moreover, twins could be considered a special case and spillover problems within twin pairs could be of particular importance.

Our child maltreatment analysis belongs to a broader problem, according to which victims of any crime more frequently than non-victims offend in the future. Different theories, mostly from the fields of sociology and criminology explain why we may observe such behavioral patterns, but the empirical findings are mixed. We explore the problem in Chapter 3 using a representative sample from the Dutch population. We focus on the victim-offender direction of the relationship and start with an OLS with a large set of controls. Next, we follow the approach by Altonji et al. (2005) and use selection on observables to gauge the selection on unobservables. We find that there are no strong short- and long-term effects of victimization on current offending.

Social science theories point that education could be a way to prevent criminal involvement. If an individual gains a lot of skills and knowledge, he will have a higher opportunity cost if selecting a criminal career than someone without sufficient educational built-up. In Chapter 4, we focus on human capital accumulation but from a different perspective. Concretely, we investigate the change of the relative age effect on test scores across the Dutch primary educational system, i.e. whether the older students within a class are doing better than their younger peers and how persistent is such a difference. We find that in grade 2 the relatively older students perform about 0.9 standard deviations better than their relatively younger peers on a math and language test. By grade 8, this gap has decreased to 0.3 standard deviations but remains significant. It is important from a policy perspective to understand why this decrease occurs. The Dutch educational system has different policy measures to target poor performing students, such as remedial classes, retention, and sending students to special education. We find that all of these contribute to the younger students catching up to their older peers over time, and in addition, there could be maturity (time) effect. Interestingly, for students from low socio-economic background the gap over grades does not decrease as much. Therefore, this group could benefit from additional attention and help.

In Chapter 5 we study what factors matter for the well-being and how it is affected by different traumatic experiences. After all, a certain negative experience is most of all interesting from a policy perspective if it has long-standing negative effect on the well-being and the quality of life. We use again a large sample of fraternal and identical twins, with an average age of about 30 who experienced a traumatic event in the near past. Our findings are in line with the evidence in the happiness literature. We establish that the marital status, the self-reported health, low income and education are important determinants for one's well-being. We also use the timing of different traumatic events that happened in the near past and find that in the first one to three years after an assault, a rape, or an accident, the self-reported emotional well-being of the respondent is substantially lower in comparison to his sibling who has not been through something this traumatic. After three years, there is less impact of the different negative experiences. Here, spillover effects within the twin pair could be less problematic because we can control for a traumatic experience of one's sibling.

Chapter 2

The Effect of Child Maltreatment on Illegal and Problematic Behavior. New evidence on the "cycle of violence" using twins data[1]

2.1 Introduction

Recent scandals with TV personalities, pop stars, and Catholic clergy have attracted massive media attention towards the problem of child maltreatment. However, these cases are only part of a much larger problem. According to The World Health Organisation, approximately 20% of women and 5–10% of men report being sexually maltreated as children, while 23% of people report physical maltreatment.[2] The consequences of childhood maltreatment might be significant, not only for the maltreated individual but also for society at large. A large literature has documented that maltreated and neglected children perform worse in school, have lower cognitive abilities, and display a number of mental problems such as depressions and difficulties to cope with everyday life. In addition, child maltreatment has been positively associated with the use of illegal drugs, anti-social behavior, and crime (e.g. Gil (1970), Kempe *et al.* (1962), Hunter & Kilstorm (1979), Dinwiddie *et al.* (2000), Nelson *et al.* (2006)).

In this paper we investigate the effect of child maltreatment on illegal and problematic

[1]This chapter is based on joint work with Prof. Dinand Webbink and Prof. Nicholas Martin. The paper benefited from suggestions of participants in the TI PhD seminars, the participants at the Augustin Cournot doctoral days in Strasbourg (2013) and the 25th EALE conference in Turin (2013).

[2]http://www.who.int/mediacentre/factsheets/fs150/en/

behavior such as drug abuse, conduct disorder, and criminal involvement in adulthood, using unique data of Australian twins. This effect is proposed in the literature about the so-called 'cycle of violence' hypothesis. This hypothesis links child maltreatment to future criminal behavior (Widom (1992), Widom (1989), Widom & Maxfield (2001), Currie & Tekin (2012)). Studying the long-term effects of child maltreatment might therefore add to the understanding of adult criminal behavior. The effect is also particularly important because both drug abuse and criminal behavior are associated with high social costs. For instance, the costs of crime for Australia, the context of this paper, have been estimated at 5% of GDP and the yearly costs of illicit drug use at $56 billion (Mayhew (2003)).

Investigating the effects of childhood maltreatment on illegal and problematic behaviors poses several empirical challenges. First, obtaining reliable measures of child maltreatment is difficult. Administrative data are likely to capture only a small proportion of the actual occurrences of childhood maltreatment. In addition, parental reports on child maltreatment might not be reliable due to the high sensitivity of the topic and reluctance of parents to report about their own or other's misbehavior. In this paper we are able to use retrospective self-reports on childhood maltreatment from a sample of twins aged 24 to 36 years. The measures are based on an extensive questionnaire that focused on various severe types of sexual and physical child maltreatment.

Second, estimates of the effect of childhood maltreatment on illegal and problematic behavior might be confounded by unobserved factors and by reverse causality. Nearly all previous studies on the long-term effects of child maltreatment use estimation strategies, such as matching or linear regression, that might suffer from omitted variable bias. However, children that are maltreated or the families in which maltreatment occurs are probably not a random draw from the population. For instance, Paxson and Waldfogel (1999, 2002) find that children in families of lower socioeconomic status have a higher probability of being maltreated or neglected. As a consequence, unobserved factors that are correlated with childhood maltreatment and the outcome variables might bias the estimated effects of maltreatment. In this paper we try to mitigate this concern by us-

ing within-family estimation. This approach might reduce the endogeneity problem as it controls for all unobserved factors that are shared by family members. In a recent paper Currie and Tekin (2012) apply this approach and especially focus on variation within siblings.[3]

In this paper we are able to apply this approach to a large sample of twins. As our data also contain information about the zygosity of the twins we can distinguish between fraternal and identical twins. The advantage of using twins instead of siblings is that the family fixed effect is expected to capture more unobserved factors. For instance, the family circumstances will typically be more similar with twins than with siblings. In addition, identical twins are genetically identical, whereas siblings on average only share half of their genetic endowments. A further empirical challenge in estimating the effect of childhood maltreatment is reverse causality. Our data also include information about the timing of the childhood maltreatment and the timing of the illegal and problematic behavior. We exploit this information to address the potential threat of reverse causality.

Although child maltreatment may have large economic and social consequences, it has hardly been studied by economists. To our knowledge, Currie and Tekin (2012) and Paxson and Waldfogel (1999, 2002) are the only studies on this topic in the economic literature. Our study contributes to the economic literature by adding a new piece of evidence. We follow the same empirical strategy as Currie and Tekin (2012; henceforth CT) but we can extend their analysis in several important ways. First, we are using a large sample of twins and we are able to distinguish between identical and fraternal twins. CT mainly focus on sibling comparisons. Their twin sample is small and they cannot control for zygosity. By using a large sample of twins and differentiating between identical and fraternal twins we can improve the empirical strategy as twins, and especially identical twins, are expected to share more unobserved factors than siblings. Second, in the present paper we are able to look at some other important outcomes new to the literature, such as drug dependence and conduct disorder, and we are able to differentiate the effect of

[3]Currie and Tekin (2012) also use a small sample of twins but their main estimates are based on a sample of siblings.

sexual abuse according to the perpetrator. Third, we can take account of reverse causality by exploiting data about the timing of child maltreatment and the timing of illegal and problematic behavior. Fourth, our study contributes to the understanding of the long-term effects of physical maltreatment as our data contain information about severe types of physical maltreatment. Finally, by looking at a different country we can also test the external validity of the important findings on the 'cycle of violence' in CT (2012).

We find consistent evidence that childhood maltreatment has a large effect on illegal and problematic behavior both between and within families. The estimated effects imply an increase of illegal and problematic behavior of fifty to one hundred percent relative to the baseline levels of individuals that have not been maltreated. This is consistent with the findings by Currie & Tekin (2012) who report a doubling of criminal behavior due to childhood maltreatment. Both physical and sexual maltreatment have a large effect on all types of illegal and problematic behavior. Sexual maltreatment by an outsider increases externalising behavior such as conduct disorder and crime. Sexual maltreatment by a family member induces internalising behavior such as drug abuse or drug dependence. For males we find that physical and sexual maltreatment increase all types of illegal and problematic behavior. For females we find that both sexual maltreatment and physical maltreatment increase drug abuse (dependence) and conduct disorder. In sum, our estimation results, based on OLS models with a large set of controls and twin fixed effect models, indicate a strong relationship between child maltreatment and illegal behavior. We acknowledge that the variation that is used for obtaining these results is not as clean as the variation that is obtained in a randomized setting. However, in the context of child maltreatment it is very difficult, and probably impossible, to obtain variation that is really exogenous. Therefore, our approach, which extends the control strategies applied in previous studies towards samples of twins and identical twins, might be considered as the best feasible approach for investigating the causal effect of child maltreatment. We believe that our results, which are consistent for different models and samples, should be interpreted as further evidence for the so-called 'cycle of violence'.

This chapter is organised in the following way. Section 2.2 discusses the theoretical background of the problem and summarizes the previous literature. Section 2.3 discusses the data, the construction of the main variables and provides descriptive statistics. In Section 2.4 we discuss the empirical strategy. Section 2.5 shows the main estimation results. Section 2.6 provides various robustness analyses and Section 2.7 concludes.

2.2 Previous studies on child maltreatment and illegal behavior

A large literature, both in economics and in other social sciences, has investigated the long-term effects of early childhood conditions. This literature studies the effects of negative shocks in early childhood, such as poverty, malnutrition, disease exposure, stress, etc. The effects of such negative shocks frequently manifest in chronic diseases later in life (Gluckman *et al.* (2008)). Early life conditions affect the infant's health, his/her educational attainment and labour market outcome. According to a model by Currie & Stabile (2003), individuals are endowed with their maximal health stock when they are born and afterwards they are exposed to various shocks which decrease it. Childhood maltreatment can be considered an important topic within this literature as maltreatment can be seen as a severe negative shock that potentially may affect various future outcomes, such as academic performance, mental health, problematic behavior, and crime.

2.2.1 Theories about the effects of child maltreatment

In the literature several theories have been formulated to explain the link between childhood maltreatment and problematic behavior. Antisocial behavior and delinquency among victims of child sexual maltreatment usually are interpreted as anger stemming from the traumatic experience and representing a desire for retaliation (Finkelhor & Browne (1985)). Childhood physical maltreatment could also lead to chronic aggressive behavior by hav-

ing an impact on the development of social-cognitive processes (Dodge & James E. Bates (1990)). According to the so-called attachment theory, physical maltreatment will prompt the child to perceive the world as a threatening place and maltreated children become hyper-vigilant towards hostile cues and respond to the behavior of others with violence. The context of maltreatment can also be linked with control theory, which states that maltreatment disrupts social bonds and the individual does not restrain from violence but gratifies his natural impulses (Benda & Corwyn (2002)). Finally, social learning theory says that being the victim of childhood maltreatment provides a model of violence the individual will follow in his/her adult life because experiencing violence leads one to evaluate aggressive behavior as leading to efficacious positive outcomes (Dodge & James E. Bates (1990)). Thus, social learning theory leads to the formulation of the cycle of violence hypothesis, according to which childhood maltreatment predisposes to violence in later years (Widom (1989)).

Criminological literature indicates that antisocial behavior tends to be fairly stable over the life course (Simons *et al.* (1995), Caspi & Moffitt (1995), Loeber (1982)). That is, antisocial behavior shows the characteristics of a behavioral trait. This is the so-called homotypic continuity (Pajer (1998)). In our context, this means that someone who displays aggressive behavior in adolescence will continue to do so in adulthood, which implies that conduct disorder might be a good predictor of criminal behavior later in life.

From an economic perspective, individuals engage in crime after weighing the costs and benefits that stem from illegal behavior and indulge in such behavior if the benefits exceed the costs (Becker (1968)). Control theory points out as a cost the broken social bonds but ignores the other costs and the benefits. Social learning theory suggests that the individual copies the violence he experiences and eventually accumulates capital as a criminal instead of as a regular worker (Currie & Tekin (2012)). The "homotypic continuity" hypothesis states that individuals adopt criminal careers throughout their lifetime, which is in line with the social learning theory.

In sum, the above theories imply that maltreated children will tend to grow up as ag-

gressive and mistrusting individuals who will either escape reality by engaging in substance abuse or will externally express their aggression (or both). Furthermore, individuals who engage in conduct disorder have a high propensity to continue committing crimes throughout their adulthood.

2.2.2 Empirical studies on the effects of child maltreatment

A number of social studies deal with both the short and long-term consequences of childhood sexual maltreatment. In a review study, Putnam (2003) confirms the significance of child sexual maltreatment as a factor for psychopathology, especially depression and substance abuse. Using a community sample of British women, Mullen *et al.* (1993) establish a positive correlation between child sexual maltreatment and a range of psychopathology measures, among which substance abuse and suicidal behavior. Dube *et al.* (2003) use a sample of adults from California to investigate the relationship between illicit drug use and adverse childhood experiences, among which were physical and sexual child maltreatment. These studies, though supporting the hypothesis of the negative effects of childhood sexual maltreatment, are based on associations which might be biased by unobserved factors that are related to childhood maltreatment and the outcomes variables. Several studies investigate the relationship between child maltreatment and drug addiction. The associated negative self-esteem or self-derogation resulting from child maltreatment might initiate self-destructive behavior, such as illicit drug use and alcohol consumption (Kaplan (1980), Dembo *et al.* (1987), Dube *et al.* (2003)). Maltreated children (both physically and sexually maltreated) are ill-equipped for meaningful relationships, they mistrust others, often are insecure, cognitively impaired and many of them use drugs as a way to escape from their pain and problems (Garbarino & Gilliam (1980), Dembo *et al.* (1987)).

There are very few studies that investigate the effect of maltreatment on delinquent behavior and further criminal involvement of individuals. One such study was conducted by Widom (1989), who proposed the above mentioned 'cycle of violence' hypothesis. Us-

ing a longitudinal study of substantiated cases of maltreatment and neglect, and matching them to non-maltreated children with the same background characteristics, she finds that childhood maltreatment and neglect increase the likelihood of being arrested as a juvenile by 53 percent and as an adult by 38 percent. This finding confirms the cycle of violence hypothesis. English *et al.* (2002) replicate these results and also find that emotional maltreatment increases the risk of later violent behavior. However, the matching techniques that are used in these studies are based on the conditional independence assumption which might not hold as child maltreatment probably is not (conditionally) random.

Several studies have used data of twins to investigate the negative consequences of child maltreatment. Nelson *et al.* (2006) use the same data as used in this paper to investigate the association between a history of childhood sexual maltreatment and the use of specific licit (nicotine and alcohol) and illicit drugs (cannabis, sedatives, cocaine). They use a survival analysis without exploiting the twin dimension of the data. They find that a history of childhood sexual maltreatment is associated with a significant risk for regular smoking and illicit drug use. Dinwiddie *et al.* (2000) investigate the association between reporting childhood sexual maltreatment and psychopathology. They find that childhood sexual maltreatment is positively associated with lifetime diagnoses of major depression, conduct disorder, panic disorder, and alcoholism. Moreover, they find that individuals that report childhood sexual maltreatment are also more likely to report suicidal ideation and a history of suicide attempts.

In the economic literature we are aware of only one paper that studies the long-term effects of child maltreatment. Currie and Tekin (2012) investigate the impact of child maltreatment on the likelihood of committing crime using a national representative survey from the US (the National Longitudinal Study of Adolescent Health). Using OLS, sibling and twin fixed effect estimation they find that maltreatment doubles the probability of engaging in crime, with sexual abuse having the largest impact. Moreover, they confirm that the probability of engaging in crime increases with the severity of maltreatment. These conclusions are mainly based on the results of the sibling fixed effect estimations.

2.3 Data

We use data from the so-called younger cohort of twins of the Australian Twin Register (ATR). This cohort consists of a sample of 4246 twin pairs born between 1964 and 1971. The twins were registered with the ATR as children by their parents in response to media appeals and systematic appeals through the school system in the period 1980-1982. The data have been collected in two surveys. The first survey was conducted by a mailed questionnaire in 1989-1990, when the twins were 18-25 years old. The response rate of this survey was 63%. The second survey was conducted by telephone interviews in 1996-2000. The telephone interviews were completed with 6267 individuals, 2805 men (889 complete and 1027 incomplete pairs) and 3462 women (1215 complete and 1032 incomplete pairs). At the time of the interview, the twins were 24-36 years old (on average 30 years). The individual response rate for the second survey was 86 %.

The surveys gathered information on the respondent's family background (parents, siblings, marital status, and children), socioeconomic status (education, employment status, and income), health behavior (body size, smoking, and drinking habits), conduct disorder, personality, feelings, and attitudes. Zygosity was determined by a combination of diagnostic questions plus blood grouping and genotyping. The measures of child maltreatment and illegal behavior were gathered in the second survey which is called the Semi-Structured Assessment for the Genetics of Alcoholism. In the analysis we focus on complete twin pairs which give us a sample of 2330 identical and 3200 fraternal individual twins. In the analyses that look at the effects for male or female pairs of twins we exclude the opposite sex twin pairs. This reduces the number of observations. Moreover, in some analyses the number of observations will be smaller due to missing values on the outcome variable.

A. Measuring child maltreatment

Various questions about sexual and physical maltreatment are asked in the survey. We

use these questions for constructing indicators of sexual and physical maltreatment.

Sexual maltreatment

As an indicator for sexual maltreatment we constructed a dummy variable that equals one when the respondent answered positively to either of the questions below:

(1) Before the age of 16 he/she was forced into any sexual contact with anyone else other than a family member (someone older by five years);

(2) Before the age of 16 he/she was forced into any sexual contact with a family member (someone older by five years);

(3) He/she was sexually molested as a child.

The data also provide information about the frequency of sexual maltreatment (once or multiple times) and whether the offender was a family member or an outsider. We also use this information in the analysis.

Physical maltreatment

The questions about physical maltreatment refer to the ages between 6 and 13. From these questions we constructed a dummy variable that equals one when the respondent replied positively to either of the statements below:

(1) He/she was often or sometimes punched or hit with a belt or stick or something like that by either of the parents or physically punished so that he/she hurt the next day;

(2) The way in which either of the parents punished him/her was harsh or the respondent was ever physically injured or hurt on purpose as a child (examples include broken bones, bruises, punishments that included scalding water or any other physical injuries).

(3) The respondent was physically abused as a child.

Like Currie and Tekin (2012) and many other studies, this paper uses retrospective reports of maltreatment provided by the individuals themselves. The self-reported retrospective nature of the data might be a reason for concern as individuals might forget past experiences. The fact that we focus on severe types of maltreatment and use various questions for measuring these types of maltreatment might mitigate this concern. In addition, the data on maltreatment were collected when the twins were still relatively young

(24-36 years old). We also checked whether twins aged below the sample mean more often reported sexual or physical maltreatment than twins aged above the sample mean. This was not the case (13 % versus 12 % for sexual maltreatment; 35 % versus 34 % for physical maltreatment). Of course, some individuals might be reluctant to give a truthful answer to the maltreatment questions due to their sensitivity. We discuss the issue further in Section 2.1.6.

Table 2.1 shows the proportions of individual twins that report some kind of maltreatment measured by the indicators described above. Approximately 12 % of the sample reports sexual maltreatment. This proportion lies in the same ball park as the proportions reported in other studies for Australia. The prevalence of sexual maltreatment for males in Australia ranges from around 10% (Mamun *et al.* (2007)) to 16% (Dunne *et al.* (2003)) and from 12 % (Dunne et al. (2003)) to 42 % (Mazza & Dennerstein (2001)) for females. Table 2.1 also shows substantial variation in reporting sexual maltreatment within families (columns (2) and (5)); in 67 % of the twin pairs that reported sexual maltreatment there is variation in reporting (52 % in the identical twins sample), and 64 (52) % of those reporting sexual maltreatment by a family member or by an outsider have a discordant report from their (identical) co-twin. Physical abuse refers to severe physical maltreatment. It is the most common maltreatment in the data set, reported by 34% of all pairs. The fraction reporting physical abuse is higher than the fraction reporting sexual abuse and it is higher than the prevalence rate of physical abuse for Australia that other studies discover (Mouzos & Makkai (2004) report a rate of physical abuse of 18%). The higher prevalence of physical maltreatment might be explained by the specific nature of our sample consisting of twins. Previous studies have found increased rates of maltreatment (and neglect) for twins (Robarge *et al.* (1982), Nelson & Martin (1985), Currie & Tekin (2012)). The higher prevalence of physical maltreatment among twins might result from the higher stress parents are under, exhaustion, financial pressures, neonatal complications, and higher rates of premature birth among twin pairs (Robarge *et al.* (1982), Nelson & Martin (1985)). We also observe substantial variation in the physical maltreat-

ment reported within families: 47% in the whole sample of physically maltreated twins (42 % of the physically maltreated identical twins) have a physical maltreatment report different from that of their co-twin. This variation within twin pairs is important for the identification of the effects of maltreatment. [4]

Table 2.1 Fraction Reporting Maltreatment in the total sample

Maltreatment	(1) All twins	(2) Pairs with different reports (%)	(3) Fraternal twins	(4) Identical twins	(5) Pairs with different reports (%)
Sexual	0.12	0.67	0.12	0.11	0.52
Sexual by outsider	0.04	0.78	0.04	0.04	0.61
Sexual by family member	0.04	0.64	0.04	0.04	0.52
Physical	0.34	0.47	0.34	0.34	0.42
Any	0.40	0.44	0.40	0.40	0.39
Observations	5530		3200	2330	

Note: The sum of those reporting sexual maltreatment by a family member and those reporting sexual maltreatment by an outsider is smaller than the total fraction because sexual maltreatment is defined by three variables.

B. Measuring illegal and problematic behavior

In the questionnaire several questions about illegal and problematic behavior have been asked. We use these questions for constructing indicators for drug abuse, drug dependence, conduct disorder, and crime.

Drug abuse

We follow the American Psychiatric Association's (APA) definition of drug abuse based on the Diagnostic and Statistics Manual of Mental Disorders (DSM-IV). As an indicator of drug abuse we use a dummy variable that equals one when the respondent answered positively to either of the statements below:

[4] The physical maltreatment is an interesting phenomenon and social science studies propose as an explanation a "single child" targeting where the parents would maltreat only one of the children in the family. This could be due to some characteristics of the child (his gender, idiosyncratic behavior, physical and mental problems), or to some parental characteristics (mental disorders, abuse of alcohol, drugs, etc.). Some researchers also argue that abusive parents might targer a single child instead of all children in the family in order not to attract much outside attention, or prevent cooperation between the children and their reporting to the authorities (see Jaffee *et al.* (2004)).

(1) He/she has been under the influence of a certain drug, which increased his/hers chances of getting hurt (examples include driving a car or a boat, using knives, machinery or guns, crossing against traffic, climbing or swimming);

(2) Being under the influence of the drug ever interfered with working, studying or taking care of household responsibilities.

Drug dependence

Our indicator for drug dependence is based on a series of questions. The indicator has a value of one when the respondent replied positively to at least two of the following questions dealing with drug dependence:

(1) He/she has ever used the respective drug(s) for more days or in larger amounts than intended;

(2) Whether compared to the first time the respondent used the respective drug(s), he/she needed increasingly larger amounts to get any effect or he/she no longer was getting high on the same amounts as before;

(3) The respective drug(s) has ever caused emotional or psychological problems, like feeling depressed or uninterested in things, feeling grumpy or easily irritated, having trouble thinking clearly for more than 24 hours, feeling paranoid or suspicious of people;

(4) Whether there have been three or more times that the respondent wanted to cut on the respective drug(s);

This definition of drug abuse includes three out of the four criteria used in the APA definition and does not completely cover the DSM-IV diagnostic criteria. However, previous studies using the same data also use this definition and show that it provides a valid measure of drug abuse and dependence (see Lynskey *et al.* (2002, 2003)).

Conduct disorder

The American Psychiatric Association (APA) defines conduct disorder as "a repetitive and persistent pattern of behavior in which the basic rights of others or major age-appropriate societal norms or rules are violated, as manifested by the presence of three (or more) of the following criteria in the past 12 months". For instance, criteria like: often

initiated physical fights; has deliberately destroyed others' property; has broken into someone else's house, building, or a car; has often been truant from school. Our data contain self-reported information on 21 statements that reflect behavioral problems before the age of 18. We used this information to construct two measures of conduct disorder. Our first measure follows the APA-definition and is based on the question 'Did you do at least 3 of these things within the same 12-month period?'. Twins who responded 'yes' were coded as 1, twins who responded 'no' were coded as 0. We call this measure the APA-definition of conduct disorder. For our second measure of conduct disorder we created a conduct disorder score based on the 21 statements.[5] We call this measure the conduct disorder score.

Crime

Our indicators of crime are based on three general questions about criminal behavior. The first indicator is coded as one in case the respondent answered positively to at least one of the following questions:

(1) He/she has ever spent time in jail;

(2) Has ever been arrested (for anything else other than drunk driving or drunken behavior);

(3) Has ever done something (else) that he/she could have been arrested for (even though he/she was not).

The last question is about crime that has remained unnoticed by the authorities. It is important to also take these crimes into account as they may generate high social costs. We call this first indicator 'total crime'. The second indicator of crime is based on the first two crime questions only. We call this 'detected crime'. It should be noted that due to the routing of the questionnaires not all individuals had to answers these 'crime' questions. Individuals that did not report any of a series of problematic behaviors related to conduct disorder did not have to answer the 'crime' questions. This implies that our indicator of crime is conditional on having at least one behavioral problem used in measuring conduct

[5] Further details about the construction of these measures can be found in Webbink *et al.* (2012)

disorder. As mentioned above, conduct disorder is defined as displaying at least three problematic behaviors within one year. Since conduct disorder and other problematic adolescent behavior is viewed in the social literature as a relatively good predictor of later crime engagement (Simons *et al.* (1995), Loeber (1982), Pajer (1998)), it might be expected that the rooting of the questionnaire will not induce substantial measurement error for the crime indicator.

These measures of illegal and problematic behavior are also based on self-reports. The reliability of these self-reported data is an important issue. In criminology the use of self-reported data is well established. Self-reported data collection has been the dominant technique used for measuring illegal behavior since its introduction in the 1950s by Short & Nye (1957). A large literature shows that self-reported data have consistently acceptable reliability and validity (see Webbink *et al.* (2013)).

Table 2.2 shows the proportion of the twins that report some type of illegal or problematic behavior. In addition, this table reports the proportions for each category of maltreatment. The first column of Table 2.2 shows that 17% of the total sample reports drug abuse. Official reports about drug prevalence in Australia (Australian Institute of Health and Welfare 2006) indicate that around 38% of Australians aged 14 and more used some illicit drug at a certain point in their life, and around 15% had used illicit drugs at least once in the past 12 months. Although these statistics refer to any drug use, which does not necessarily imply drug abuse and/or dependence, the prevalence rates in our sample do not seem implausibly high or low compared to these statistics. In the total sample we find a prevalence rate of conduct disorder of 13%. This is in line with Searight *et al.* (2001) who report that approximately 6 to 16 percent of boys and 2 to 9 percent of girls meet the diagnostic criteria for conduct disorder. In the total sample 27% of the twins report positively about one of the three crime questions. The answers on the first two crime questions reported in our sample seem consistent with population statistics (Webbink *et al.* (2013)).

Table 2.2 also shows the proportions that report some type of illegal or problematic

behavior within each category of maltreatment. We observe that drug abuse is much more prevalent among those that report any or a specific type of maltreatment (columns (3) to (7)) than among those who report no maltreatment (column (2)). For instance, the prevalence of drug abuse among those who report no maltreatment is 13 percent against 23 percent among those who report any maltreatment. A similar pattern can be observed for the other types of illegal behavior; the prevalence of illegal behavior is substantially higher among those that report maltreatment than among those that do not report maltreatment. In addition, it can be observed that the prevalence of any illegal behavior is 16 percentage points higher among those that have been maltreated versus those that have not been maltreated from a baseline of 31 percent. These statistics suggest a link between maltreatment and illegal and problematic behavior. However, these statistics might provide a biased picture because the occurrence of maltreatment is probably not random.

Table 2.2 Means of outcome variables by maltreatment and type of maltreatment

		Maltreatment					
	(1) Full sample	(2) None	(3) Any	(4) Physical	(5) Sexual	(6) Sexual by family	(7) Sexual by outsider
Drug abuse	0.17	0.13	0.23	0.23	0.29	0.27	0.25
	[0.37]	[0.33]	[0.42]	[0.42]	[0.45]	[0.44]	[0.48]
Drug dependence	0.15	0.11	0.21	0.21	0.28	0.20	0.20
	[0.36]	[0.32]	[0.40]	[0.41]	[0.45]	[0.46]	[0.46]
Conduct disorder:							
APA definition	0.13	0.08	0.19	0.22	0.22	0.10	0.11
	[0.33]	[0.27]	[0.39]	[0.41]	[0.42]	[0.30]	[0.31]
Score	1.79	1.32	2.49	2.58	2.79	2.70	2.93
	[2.39]	[1.95]	[2.78]	[2.86]	[2.87]	[2.79]	[2.94]
Crime:							
Total	0.27	0.21	0.35	0.37	0.32	0.31	0.27
	[0.44]	[0.41]	[0.48]	[0.48]	[0.47]	[0.46]	[0.44]
Detected	0.05	0.03	0.08	0.09	0.09	0.08	0.12
	[0.22]	[0.18]	[028]	[0.28]	[0.28]	[0.27]	[0.33]
Any illegal behavior	0.37	0.31	0.47	0.49	0.45	0.39	0.37
	[0.48]	[0.46]	[0.50]	[0.50]	[0.50]	[0.49]	[0.48]
Observations	5530	3322	2208	1894	653	227	206

Note: Standard deviations are given in brackets.

Control variables

The data provide information on various individual characteristics which can be used as control variables in the OLS models. These variables also provide insight in the type of families in which the occurrence of child maltreatment is more likely. Information is available about age and gender of the respondent. We also have information about the educational attainment (years of schooling) and about the use of alcohol of both parents. This could be important since the social science literature relates child maltreatment with both alcoholism of the parents and with their level of school attainment (McLaughlin *et al.* (2000)). We can also account for whether the child has been raised by his natural parents and whether an adoptive/step-parent was present (the presence of a step-parent is often associated with an increased risk of maltreatment). Another variable of interest is whether the child has witnessed conflict in the family (did the parents fight in front of the children). Moreover, we include control variables for the age of the mother (teenage mother, 18-30 years, 30-40 years and above 40).

Table 2.3 shows the means of the explanatory variables by the different maltreatment categories. A comparison of individuals that report some type of maltreatment with individuals that do not report maltreatment reveals that maltreated individuals are less often raised by both natural parents, more often had an adoptive or step parents, more often saw their parents fighting, more often report that mom or dad were alcoholic and more often had a teen mother. Hence, these statistics suggest that child maltreatment is more likely in families that seem to have more problems.

Table 2.3 Means of control variables for those reporting physical, sexual or no maltreatment

	All twins			Identical twins		
Maltreatment:	Physical	Sexual	None	Physical	Sexual	None
Male (%)	0.53 [0.50]	0.35 [0.48]	0.40 [0.49]	0.50 [0.50]	0.34 [0.47]	0.37 [0.48]
Mother has a high school degree (%)	0.26 [0.44]	0.26 [0.44]	0.27 [0.44]	0.26 [0.44]	0.23 [0.42]	0.26 [0.44]
Mother has more than a high school degree (%)	0.17 [0.28]	0.22 [0.41]	0.21 [0.41]	0.16 [0.37]	0.23 [0.42]	0.21 [0.41]
Father has a high school degree (%)	0.20 [0.40]	0.20 [0.40]	0.21 [0.41]	0.21 [0.41]	0.24 [0.43]	0.21 [0.40]
Father has more than a high school degree (%)	0.24 [0.43]	0.27 [0.45]	0.27 [0.44]	0.24 [0.43]	0.27 [0.44]	0.27 [0.44]
Raised by natural parents (%)	0.78 [0.42]	0.73 [0.45]	0.84 [0.36]	0.76 [0.43]	0.71 [0.46]	0.83 [0.37]
Adoptive/step parent (%)	0.09 [0.28]	0.11 [0.31]	0.07 [0.25]	0.10 [0.30]	0.13 [0.34]	0.08 [0.27]
Parents fought in front of children (%)	0.42 [0.49]	0.39 [0.49]	0.26 [0.44]	0.41 [0.49]	0.42 [0.49]	0.27 [0.44]
Respondent said mother had problems with alcohol (%)	0.04 [0.20]	0.05 [0.21]	0.02 [0.14]	0.04 [0.18]	0.04 [0.20]	0.02 [0.14]
Respondent said father had problems with alcohol (%)	0.23 [0.42]	0.23 [0.42]	0.13 [0.34]	0.21 [0.40]	0.25 [0.44]	0.12 [0.33]
Mother gave birth as a teenager (%)	0.014 [0.12]	0.01 [0.10]	0.006 [0.08]	0.018 [0.13]	0.013 [0.11]	0.005 [0.07]
Mom's age at birth 18-30 (%)	0.59 [0.49]	0.59 [0.49]	0.56 [0.50]	0.67 [0.47]	0.67 [0.47]	0.60 [0.49]
Mom's age at birth 30-40 (%)	0.24 [0.42]	0.23 [0.42]	0.28 [0.45]	0.18 [0.38]	0.18 [0.38]	0.26 [0.44]
Mom's age at birth>40 (%)	0.08 [0.27]	0.10 [0.30]	0.08 [0.27]	0.07 [0.26]	0.08 [0.28]	0.07 [0.25]

Note: Standard deviations are given in brackets

2.4 Empirical strategy

For investigating the effect of child maltreatment on illegal and problematic behavior we start by estimating OLS regressions without exploiting the twin-dimension of our data. The OLS models that we estimate have the following form:

$$Y_i = \alpha_0 + \alpha_1 M_i + \alpha_2 X_i' + \varepsilon_i \tag{2.1}$$

where Y_i is a dummy variable for illegal and problematic behavior (drug abuse, drug dependence, conduct disorder, or criminal involvement) of individual i, X_i is a vector of control variables, M_i is a dummy variable for (a type of) maltreatment and ε_i is the error term. Estimation of Equation (2.1) will yield the causal effect of maltreatment on illegal and problematic behavior if maltreatment is not correlated with unobserved factors that are also important for illegal behavior. However, this conditional independence assumption might not hold as the occurrence of child maltreatment is probably not random. For instance, previous studies indicate that child maltreatment is related to observable characteristics of families (Paxson & Waldfogel (1999, 2002)). Including a large set of controls in the model of Equation (2.1) might help but unobserved factors can still be important determinants of child maltreatment and illegal or problematic behavior. Random variation in maltreatment can solve this problem but, for obvious reasons, it seems not feasible to find variation in maltreatment induced by controlled or natural experiments. Therefore, exploiting variation in maltreatment within families might be the best feasible solution for identifying the causal effect of maltreatment on illegal behavior. Models that exploit variation within families control for all observed and unobserved factors within families that are shared by the siblings or twins. In line with Currie and Tekin (2012) we use variation in maltreatment within families. Currie and Tekin (2012) are able to use variation within sibling and within twins of which the zygosity is unknown. Our data also allow us to distinguish between fraternal and identical twins. As a second step in our analysis we will estimate models that include family fixed effects of the following form:

$$Y_{ij} = \beta_0 + \beta_1 M_{ij} + \beta_2 X_{ij} + \mu_j + \varepsilon_{ij} \tag{2.2}$$

where the index j refers to family j, and μ_j is the unobserved family fixed effect. This family fixed effect captures all factors that are shared by both twins. Fraternal twins share, just like siblings, approximately half of their genes. Identical twins have exactly the same genes. In addition, most twins are raised within the same family and will share many components of this social environment. Estimation of Equation (2.2) reduces the problem of omitted variables bias by differencing out all observed and unobserved factors shared by both twins. We will estimate models for the total sample of twins and separately for samples of identical twins only. The latter sample might be preferred as identical twins share all genes. However, focusing on identical twins only also reduces the sample size and the variation in child maltreatment that can be used in the estimation.

Although the within-twin estimator controls for all unobserved genetic and family factors that are shared by the twins, there are several concerns with this approach. The first concern is reverse causality; does child maltreatment lead to illegal behavior or does early illegal behavior induce childhood maltreatment? To address this issue we exploit information about the timing of child maltreatment and the timing of the various types of illegal behavior. To reduce the probability that our estimates will be inflicted by potential reverse causality, we re-estimate the main models after excluding all observations for which the illegal or problematic behavior might have preceded the reported maltreatment.

The second concern is measurement error as pointed out by Griliches (1979) and Bound & Solon (1999)). The within-family estimator exacerbates measurement error, which is likely to bias the estimates towards zero. In Section 2.3 we already discussed the potential for misreporting due to forgetting. We further investigate the potential retrospective bias in the responses by looking at the reporting errors in questions for which we have two reports. In our data the twins were asked to report about the educational attainment of themselves and the educational attainment of their co-twin. This provides us with a measure of misreporting and we compare this measure with the findings from other

studies. We use this measure to calculate the potential downward bias of our twin-fixed effect estimator.

The third concern is spillover effects within pairs of twins. As twins grow up in the same family and might share a lot of time together they might also influence each other's behavior. For instance, if the co-twin copies the illegal behavior of the twin that experienced childhood maltreatment we will underestimate the effect of maltreatment. It is also possible that the behavior of one twin restrains the other twin from doing a specific type of behavior. This might bias the estimated effects of maltreatment. We investigate the potential bias of these spillover effects by excluding twins from our estimation sample that reported to be very close with their co-twin. Spillover effects will probably be less important for the sample of twins that do not report to be very close with their co-twin.

The fourth concern with within family estimates is endogeneity bias. Although twins, and especially identical twins, share many observed and unobserved factors they probably are not exactly identical. Hence, unobserved factors within pairs of twins might bias the estimated effects. The bias in the within-family estimator may not always be smaller than the bias in the cross-sectional estimator (Bound & Solon (1999)). This depends on the importance of the fixed family component in the unobservable factors that both affect child maltreatment and the outcome variable. This concern, which is typical for all sibling and twin studies that do not exploit exogenous variation within families, is important for the interpretation of the estimates. For establishing a causal effect of child maltreatment on illegal and problematic behavior we should assume that unobservable factors within families don't bias the estimates. This assumption might be difficult to defend as the variation in child maltreatment within pairs of twins is not as clean as the variation that is obtained in a randomized setting. This implies that a causal interpretation of our estimates might not be fully justified. The aim of our empirical strategy is to extend (and improve) the control strategies applied in previous studies by taking account of fixed effects within pairs of twins and within pairs of identical twins. Using twin fixed effects allows us to control for all genetic factors and for all environmental factors shared by (identical)

twins. As it is very difficult, and probably impossible, to obtain random variation in maltreatment this might be considered as the best feasible approach for investigating the causal effect of child maltreatment. Although this strategy has potential limitations we believe that the estimates provide important evidence about the long-term effects of childhood maltreatment and about the so-called 'cycle of violence'.

2.5 Main estimation results

In this section we report the main estimation results. We start by investigating the effect of any maltreatment on three types of illegal and problematic behavior. Next, we investigate the effect of specific types of maltreatment (sexual or physical maltreatment) on illegal and problematic behavior.

2.5.1 The effect of any maltreatment on illegal and problematic behavior

This section presents the main estimation results of the effect of any child maltreatment on illegal or problematic behavior, using both OLS and twin-fixed effect models. The main independent variable in these models is 'any maltreatment' which measures whether an individual has experienced sexual or physical maltreatment. We estimate the effect of any maltreatment on six dependent variables: drug abuse, drug dependence, two measures of conduct disorder (APA-definition and conduct disorder score) and two indicators of crime (total and detected crime). Table 2.4 shows the estimation results, each cell shows the estimate of a separate regression. Column (1) shows the results of an OLS-regression of a type of illegal behavior on any maltreatment without any controls, and column (2) – with full set of controls. All models with covariates include controls for age, gender, parental education, age of the mother at the time of birth, parental alcohol abuse, whether the individual was raised by his/her natural parents, the presence of a step/adoptive parent,

and parental fights observed during childhood. The standard errors are corrected for clustering at the twin pair level. Columns (3) to (5) show the fixed effect estimates for the sample of all twins, the sample of fraternal twins and for the sample of identical twins, respectively. Robust standard errors are shown in brackets.

Table 2.4 Estimates Of The Effect Of Any Maltreatment On The Outcome Variables

			FE		
	(1)	(2)	(3)	(4)	(5)
	OLS no controls	OLS	All twins	Fraternal twins	Identical twins
Drug abuse	0.104***	0.080***	0.054**	0.057**	0.047**
	[0.010]	[0.011]	[0.014]	[0.019]	[0.021]
Observations	5530	5530	5530	3220	2330
Drug dependence	0.102***	0.080***	0.062**	0.075***	0.041**
	[0.010]	[0.011]	[0.014]	[0.018]	[0.020]
Observations	5530	5530	5530	3220	2330
Conduct disorder:					
APA definition	0.129***	0.110***	0.067***	0.074***	0.053**
	[0.009]	[0.010]	[0.013]	[0.018]	[0.020]
Observations	5530	5530	5530	3220	2330
Score	1.168***	0.980***	0.535***	0.658***	0.313**
	[0.064]	[0.070]	[0.082]	[0.117]	[0.104]
Observations	5530	5530	5530	3220	2330
Crime:					
Total	0.158***	0.152***	0.144***	0.181***	0.081
	[0.021]	[0.021]	[0.032]	[0.041]	[0.051]
Observations	2254	2254	2254	1336	918
Detected	0.065***	0.055***	0.017	0.028	−0.003
	[0.013]	[0.013]	[0.020]	[0.027]	[0.030]
Observations	2254	2254	2254	1336	918

Note: Standard errors are given in brackets; Column (2) controls for gender, age, mother/father have only a high school degree, mother/father have more than a high school degree, raised by both natural parents, adoptive/ step parent present, parents fought in front of children, respondent said mother/father had problems with alcohol, mother gave birth as a teenager, mother's age at birth was between 18-30, 30-40 or above 40. Columns (3)-(5) include twin-pair fixed effects.
* indicates statistically significant at the 10% level, ** at the 5% level, and *** at 1% level.

The OLS-estimates in the first two columns of Table 2.4 suggest a strong association between maltreatment and illegal or problematic behavior. Maltreatment is associated with an increase in drug abuse or drug dependence of 8 (model with controls in column (2)) to 10 percentage points (model with no controls, column (1)). In addition, maltreatment is associated with an increase in the occurrence of conduct disorder as defined by APA of 11 percentage points (13 according to the model in column (1)) and an increase in the number of conduct disorder behaviors with one. Moreover, maltreatment is associated

with an increase in total crime of 15 percentage points and an increase in detected crime of 6 percentage points. Compared to the sample means in the first column of Table 2.2 these increases are substantial. Although these models include a large set of controls the estimates might be biased by unobserved factors. The estimates in columns (3) to (5) control for all factors that are fixed within twin pairs. For the first four outcomes we observe that child maltreatment has a statistically significant effect. The estimated effects are somewhat smaller than the OLS-estimates but remain substantial. Even within pairs of identical twins we find that child maltreatment increases drug abuse, drug dependence, and conduct disorder. We also find a statistically significant and large effect of child maltreatment on total crime. For the sample of identical twins we find a positive point estimate but the estimated effect is no longer statistically significant. This might be explained by the strong reduction of the sample that could be used for this estimation because the crime outcome is not measured for individuals that report negative on all types of problematic behavior related to conduct disorder (Section 2.3) and the focus on identical twins only. In sum, the estimated effects are large. Compared to the sample means (column (1) of Table 2.2) the fixed effects estimates indicate an increase of illegal and problematic behavior with one third to two thirds. The OLS-estimates are even larger.

Table 2.5 shows estimation results obtained from separate samples of males and females. The estimation samples are smaller than in Table 2.4 because opposite sex pairs are not included.

Table 2.5 Estimates of the effect of any maltreatment by gender

Panel A. Estimates for males

		FE		
	(1) OLS	(2) All twins	(3) Fraternal twins	(4) Identical twins
Drug abuse	0.084***	0.066**	0.040	0.090**
	[0.022]	[0.026]	[0.042]	[0.032]
Observations	1778	1778	826	952
Drug dependence	0.085***	0.079**	0.076*	0.089**
	[0.020]	[0.025]	[0.039]	[0.032]
Observations	1778	1778	826	952
Conduct disorder:				
APA definition	0.133***	0.066**	0.040	0.087**
	[0.020]	[0.027]	[0.040]	[0.037]
Observations	1778	1778	826	952
Score	1.118***	0.498***	0.546**	0.437**
	[0.140]	[0.154]	[0.255]	[0.183]
Observations	1778	1778	826	952
Crime:				
Total	0.121***	0.117**	0.093	0.155**
	[0.031]	[0.051]	[0.076]	[0.070]
Observations	946	946	458	488
Detected	0.051**	−0.024	−0.005	−0.045
	[0.024]	[0.035]	[0.054]	[0.044]
Observations	946	946	458	488

Panel B. Estimates for females

		FE		
	(1) OLS	(2) All twins	(3) Fraternal twins	(4) Identical twins
Drug abuse	0.079***	0.032	0.056*	0.013
	[0.015]	[0.020]	[0.030]	[0.027]
Observations	2430	2430	1052	1378
Drug dependence	0.083***	0.033*	0.057	0.011
	[0.015]	[0.019]	[0.030]	[0.025]
Observations	2430	2430	1052	1378
Conduct disorder:				
APA definition	0.099***	0.050**	0.075**	0.023
	[0.014]	[0.016]	[0.025]	[0.022]
Observations	2430	2430	1052	1378
Score	0.943***	0.430***	0.617***	0.251**
	[0.092]	[0.101]	[0.172]	[0.119]
Observations	2430	2430	1052	1378
Crime:				
Total	0.194***	0.142**	0.254***	0.019
	[0.037]	[0.052]	[0.077]	[0.075]
Observations	772	772	342	430
Detected	0.038**	0.050*	0.036	0.076**
	[0.017]	[0.026]	[0.037]	[0.038]
Observations	772	772	342	430

Note: Standard errors are given in brackets; control variables included (see Table 2.4);
* indicates statistically significant at 10% level, ** at 5% and *** at 1%

For males we find large effects of child maltreatment on the first five outcomes. Even within pairs of identical twins we find that child maltreatment has a large and statistically significant effect on these five outcomes. For women we find a similar pattern but the estimated effects are smaller and not always statistically significant, especially in the sample of identical twins.

2.5.2 The effect of different types of maltreatment on illegal and problematic behavior

In Table 2.6 we proceed with investigating the effect of different types of maltreatment on each type of illegal or problematic behavior. Each row shows the estimated effects of a specific type of maltreatment (physical, sexual, sexual by offender, and sexual by frequency) on the six outcomes. All models include a twin fixed effect, and are separately estimated for the sample of all twins and for the sample of identical twins only (MZ).

The first row in Table 2.6 shows the effect of physical maltreatment on the three types of illegal or problematic behavior. The estimates show that physical maltreatment increases each type of illegal or problematic behavior. The estimated effects are statistically significant both for the sample of all twins as for the sample of identical twins only. The size of the estimates is slightly smaller when we use variation within pairs of identical twins. Only for the second indicator of crime we find no effect of physical maltreatment. The point estimates for the other five outcomes suggest large effects relative to the sample mean. For instance, physical maltreatment increases drug abuse or drug dependence with more than a half, it approximately doubles conduct disorder as defined by the APA and increases total crime with 50%. The second row in Table 2.6 shows the estimated effects of sexual maltreatment. For the sample of all twins we observe that sexual maltreatment has a large and statistically significant effect on all indicators of the three types of illegal or problematic behavior. The estimates for the sample of identical twins are less precise but indicate that sexual maltreatment increases the three types of behavior. Within pairs

of identical twins we also find a statistically significant effect on crime detected by the authorities.

The data allow us to distinguish between the types of sexual maltreatment. The third and fourth rows show the estimated effect of sexual maltreatment by a family member or sexual maltreatment by an outsider. We observe that sexual maltreatment by a family member increases drug dependence for the sample of all twins and also for the sample of identical twins only, but does not have an effect on criminal behavior. Remarkably, sexual maltreatment by an outsider affects the two other types of behavior (conduct disorder and crime). The difference in findings between the third and fourth row suggest that the perpetrator is important for the long-term effects. An outside perpetrator seems to increase externalizing behavior whereas maltreatment by a family member increases internalizing behavior. The last two rows of Table 2.6 use data on the frequency of sexual maltreatment. We don't observe a clear pattern for the effects of sexual maltreatment that occurred once. Multiple sexual maltreatment increases drug abuse and dependence, and also increases conduct disorder. This pattern is quite similar to the estimated result from sexual maltreatment by a family member.

Next, we investigate whether the effects of specific types of maltreatment differ by gender. Table 2.7 shows the estimated effect for physical and sexual maltreatment for males (Panel A) and females (Panel B). Sample size limitations do not permit to further differentiate between sexual maltreatment by perpetrator or between one-time or multiple-times one. The most prominent difference in the estimates in Panel A and Panel B is related to the type of illegal or problematic behavior. For males we find that physical maltreatment and sexual maltreatment increase all three types of illegal or problematic behavior. The effects are also found within pairs of identical twins. For women we especially find effects on drug abuse, drug dependence, and conduct disorder. The effects on crime are less clear, which could also be due to the overall lower crime prevalence among women.

Table 2.6 Estimates of the effect of type of maltreatment

	Drug abuse		Drug dependence		CD APA definition		CD score		Total crime		Detected	
	(1) All twins	(2) MZ twins	(3) All twins	(4) MZ twins	(5) All twins	(6) MZ twins	(7) All twins	(8) MZ twins	(9) All twins	(10) MZ twins	(11) All twins	(12) MZ twins
Physical maltreament	0.048**	0.043**	0.060***	0.053**	0.067***	0.047**	0.517***	0.296**	0.116***	0.087*	−0.003	−0.013
	[0.015]	[0.022]	[0.014]	[0.021]	[0.014]	[0.021]	[0.086]	[0.109]	[0.032]	[0.051]	[0.020]	[0.030]
Sexual maltreatment	0.094***	0.053	0.078***	0.031	0.084***	0.106**	0.684**	0.428**	0.096**	0.010	0.045*	0.085**
	[0.021]	[0.033]	[0.020]	[0.032]	[0.020]	[0.033]	[0.122]	[0.169]	[0.043]	[0.072]	[0.026]	[0.042]
By family member	0.067*	0.008	0.109**	0.125**	0.044	0.034	0.200	0.279	−0.009	−0.041	−0.033	0.003
	[0.037]	[0.058]	[0.035]	[0.055]	[0.035]	[0.057]	[0.213]	[0.293]	[0.068]	[0.110]	[0.042]	[0.064]
By outsider	0.092**	0.018	0.056*	0.004	0.079**	0.114*	0.706***	0.371	0.132*	0.085	0.106**	0.200**
	[0.035]	[0.054]	[0.033]	[0.051]	[0.033]	[0.053]	[0.200]	[0.272]	[0.068]	[0.104]	[0.042]	[0.061]
One-time	0.055	−0.011	0.043	0.033	0.016	0.041	0.763**	0.417	0.047	0.156	0.077	0.071
	[0.041]	[0.063]	[0.039]	[0.060]	[0.039]	[0.062]	[0.237]	[0.320]	[0.077]	[0.131]	[0.047]	[0.077]
Multiple times	0.102**	0.066	0.105**	0.077	0.087**	0.085*	0.268	0.187	0.030	−0.084	−0.009	0.078
	[0.033]	[0.051]	[0.031]	[0.048]	[0.031]	[0.050]	[0.187]	[0.256]	[0.061]	[0.096]	[0.038]	[0.056]
Observations	5530	2330	5530	2330	5530	2330	5530	2330	2254	918	2254	918

Note: CD stands for conduct disorder, MZ is monozygotic twins; Standard errors are given in brackets; control variables included (see Table 2.4);

* indicates statistically significant at 10% level, ** at 5% and *** at 1%

Table 7. Estimates of the effect of physical and sexual maltreatment by gender

Panel A. Estimates for males

	Drug abuse		**Drug dependence**		**CD APA definition**		**CD score**		**Total crime**		**Detected**	
	(1) **All twins**	(2) **MZ twins**	(3) **All twins**	(4) **MZ twins**	(5) **All twins**	(6) **MZ twins**	(7) **All twins**	(8) **MZ twins**	(9) **All twins**	(10) **MZ twins**	(11) **All twins**	(12) **MZ twins**
Physical maltreatment	0.066**	0.084**	0.069**	0.068**	0.073**	0.070*	0.505**	0.395**	0.110**	0.127*	−0.036	−0.039
	[0.027]	[0.033]	[0.026]	[0.033]	[0.027]	[0.037]	[0.156]	[0.185]	[0.051]	[0.069]	[0.035]	[0.043]
Observations	1778	952	1778	952	1778	952	1778	952	946	488	946	488
Sexual maltreatment	0.151**	0.050	0.176***	0.173**	0.085	0.174**	0.861**	0.419	0.146	0.174	0.199***	0.231**
	[0.052]	[0.075]	[0.050]	[0.073]	[0.054]	[0.084]	[0.305]	[0.417]	[0.093]	[0.142]	[0.062]	[0.087]
Observations	1778	952	1778	952	1778	952	1778	952	946	488	946	488

Panel B. Estimates for females

	Drug abuse		**Drug dependence**		**CD APA definition**		**CD score**		**Total crime**		**Detected**	
	(1) **All twins**	(2) **MZ twins**	(3) **All twins**	(4) **MZ twins**	(5) **All twins**	(6) **MZ twins**	(7) **All twins**	(8) **MZ twins**	(9) **All twins**	(10) **MZ twins**	(11) **All twins**	(12) **MZ twins**
Physical maltreatment	0.032	0.002	0.042**	0.048*	0.053**	0.024	0.415***	0.240*	0.139**	0.061	0.013	0.036
	[0.021]	[0.029]	[0.021]	[0.027]	[0.018]	[0.024]	[0.109]	[0.139]	[0.053]	[0.078]	[0.026]	[0.040]
Observations	2430	1378	2430	1378	2430	1378	2430	1378	772	430	772	430
Sexual maltreatment	0.067**	0.062*	0.045*	−0.002	0.097***	0.088**	0.577***	0.468**	0.039	−0.028	0.002	0.029
	[0.026]	[0.036]	[0.025]	[0.034]	[0.022]	[0.030]	[0.134]	[0.162]	[0.060]	[0.082]	[0.030]	[0.042]
Observations	2430	1378	2430	1378	2430	1378	2430	1378	772	430	772	430

Note: CD stands for conduct disorder, MZ is monozygotic twins; Standard errors are given in brackets;

* indicates statistically significant at 10% level, ** at 5% and *** at 1%; Control variables included (see Table 2.4)

In sum, we find consistent evidence that childhood maltreatment has a large effect on illegal and problematic behavior both between and within families. Both physical and sexual maltreatment have a large effect on all three types of illegal or problematic behavior. Sexual maltreatment by an outsider increases externalizing behavior whereas sexual maltreatment by a family member increases internalizing behaviors such as drug abuse or drug dependence. For males we find that physical and sexual maltreatment increase all types of illegal behavior. For females we find that both sexual maltreatment and physical maltreatment increase drug abuse (dependence) and conduct disorder.

2.6 Robustness analysis

In this section we investigate several issues that may threaten the validity of our results. First, we investigate the issue of reverse causality. Next, we investigate the potential bias due to measurement error. Finally, we address the issue of spillovers within twin pairs.

2.6.1 Reverse causality

An important concern for our previous estimates is reverse causality. Our estimates might be biased if early illegal or problematic behavior induces childhood maltreatment. To address this issue we use information about the timing of child maltreatment and the timing of the various types of illegal or problematic behavior. We re-estimated the main models of Table 2.4 after excluding all observations for which the illegal or problematic behavior might have preceded the reported maltreatment.[6] For the analysis using drug abuse or drug dependence as outcome variable we excluded all individuals for which the onset of drug use might have taken place before the first reported act of maltreatment. More specifically, we excluded all individuals that used any type of drugs in the age

[6] Unfortunately, the information on the timing of behaviors and maltreatment is not detailed enough to allow a straightforward regression of maltreatment on early illegal or problematic behavior.

category of 6 to 13 years and those for which the sexual maltreatment took place before the onset of drug use. Note that we exclude individuals who report any drug use which does not imply drug dependence or drug abuse. This implies a very large reduction of our sample and probably is too restrictive. For the analysis of the effect on conduct disorder we excluded all individuals that reported displaying conduct disorder before the age of 13 since we know the physical maltreatment took place between the ages of 6 and 13.[7]

Table 2.8 Estimates of the effect of any maltreatment accounting for reverse causality

			FE	
	(1) OLS	(2) All twins	(3) Fraternal twins	(4) Identical twins
Drug abuse	0.131***	0.187**	0.206**	0.154
	[0.037]	[0.055]	[0.071]	[0.094]
Observations	726	726	376	350
Drug dependence	0.094**	0.156**	0.160**	0.158*
	[0.037]	[0.056]	[0.075]	[0.090]
Observations	726	726	376	350
Conduct disorder:				
APA definition	0.098***	0.065***	0.073***	0.049**
	[0.010]	[0.014]	[0.018]	[0.020]
Observations	4700	4700	2692	2008
Score	0.940***	0.568***	0.707***	0.321**
	[0.072]	[0.085]	[0.126]	[0.103]
Observations	4700	4700	2692	2008

Note: Individuals for whom the illegal or problematic behavior might have preceded maltreatment have been excluded; Standard errors are given in brackets; control variables included (see Table 2.4); * indicates statistically significant at 10% level, ** at 5% and *** at 1%

Table 2.8 shows the estimation results after excluding individuals for which reversed causation might have happened. We don't show the result on the crime outcomes because for all individuals the crimes reported occurred after the maltreatment. Hence, the results can be found in Table 2.4. The main pattern of findings is quite similar to the results in Table 2.4. Child maltreatment increases both types of illegal or problematic behavior. These effects are found between families but also within twin pairs. The estimates using pairs of identical twins are less precise which might be explained by the smaller samples that can be used in the estimation. These estimates suggest that it is unlikely that the

[7] Of course, maltreatment is endogenous and there might be some risky-seeking behavior we do not measure and cannot control for, which could be associated with maltreatment. As long as our outcomes are concerned, however, we exclude all cases where the maltreatment might have taken place after the onset of the problematic behavior.

previous results are driven by reverse causality.

2.6.2 Measurement error

Retrospective reports might induce measurement error because individuals might forget their past experience. In addition, reporting on sexual maltreatment might be especially prone to measurement error because of the sensitive nature of this subject. Individuals may feel embarrassed to report about their true experiences or it is possible that painful experiences have been subconsciously repressed. A well-known strategy to deal with measurement error in twin studies is to use a second independent measure obtained from answers of the co-twin as an instrumental variable (Ashenfelter & Krueger (1994)). Unfortunately, in our questionnaire twins were only asked about their own experiences with childhood maltreatment and not about the experiences of their co-twin. However, we may get some insight about measurement error by exploiting questions where twins were asked to provide answers about themselves and about their co-twins. In the questionnaire twins were asked about their own and their co-twin's educational attainment. This enables us to compare to what extent the self-reported educational attainment coincides with the one reported by their co-twin. This comparison of answers about educational attainment can yield some insight in the bias due to misreporting. The discordance in the reporting on educational attainment might be considered as a proxy for potential measurement error in childhood sexual maltreatment. In the total sample approximately 70% provided the same answer about the education of their co-twin as the co-twin him/herself and only 6% provided an answer very different from the one reported by their co-twin (difference of at least 3.5 years). Interpreting the discordance in reports of education as a proxy of the misreporting error implies that approximately 30% have misreported their sexual maltreatment experiences. In fact, several studies that deal with the reliability ratio of self-report of sexual maltreatment find ratios close to 0.7. Nelson *et al.* (2010) establishes a reliability ratio of around 0.8. Williams (1994) re-interviewed clinically documented cases

of sexually maltreated females 17 years after the experience had been first documented and finds that 62% confirm having been sexually maltreated. Therefore, in our analysis we will use a reliability ratio of self-reported sexual abuse of 0.7. Griliches (1979) has pointed out that within-family estimation increases measurement error by $1/(1-\rho_c)$ where ρ_c is the intra-class correlation. The intra-class correlation in the sexual maltreatment in our case is approximately 0.24 and the reliability ratio in the sexual maltreatment report is 0.7. This implies that the bias of the OLS estimator is -0.3*β. Then the bias in the twin-fixed effects estimator is $[-0.3/(1-0.24)]*\beta = -0.39*\beta$. Thus, we might expect a larger downward bias of the fixed-effect estimator compared to the OLS-estimator but the difference is not very large. This implies that the estimated effects in the previous sections are probably lower bounds of the true effects of child maltreatment.

2.6.3 Spillover effects within pairs of twins

If twins have a strong effect on each other this might bias the estimates. For instance, if the co-twin copies the illegal behavior of the twin that experienced childhood maltreatment we will underestimate the effect of maltreatment. It is also possible that the behavior of one twin restrains the other twin from doing a specific type of behavior. It is very difficult to investigate the importance of these potential spillover effects. We investigate the importance of this issue by exploiting information about 'the closeness of the twins'. Twins were asked about how often they see their co-twin. We re-estimated the main models after excluding twins that report to be very close with their co-twin. Hence, we keep in our sample only individuals who really do not see or contact each other very often.[8] This implies a strong reduction of the sample size as most twins report seeing each other quite regularly. Table 2.9 shows the estimation results for these restricted samples. Panel A shows the results for twins that report seeing each other 'once or twice a month' or less.

[8] We need to acknowledge that the frequency of contact between the siblings could be endogenous. One of the twins could have preference for risky behavior, which might make his/her sibling keep less frequent contact with them. Alternatively, the non-risk-seeking twins might want to help his sibling with a delinquent behavior and increase the contact with them.

Panel B shows the results for twins that report seeing each other 'a few times a year' or less.

Table 2.9 Fixed- effect estimates for the effect of maltreatment for samples of twins who do not have frequent contact with each other

	(1) OLS	(2) All twins	(3) Fraternal twins	(4) Identical twins
Panel A. Twins see each other once/twice a month				
Drug abuse	0.075**	0.067**	0.055	0.178**
	[0.023]	[0.029]	[0.034]	[0.055]
Observations	1466	1466	1114	352
Drug dependence	0.089***	0.079**	0.102**	0.045
	[0.022]	[0.027]	[0.032]	[0.053]
Observations	1466	1466	1114	352
Conduct disorder:				
APA definition	0.135***	0.076**	0.077**	0.055
	[0.021]	[0.028]	[0.032]	[0.063]
Observations	1466	1466	1114	352
Score	1.172***	0.630***	0.814***	0.084
	[0.151]	[0.178]	[0.211]	[0.330]
Observations	1466	1466	1114	352
Crime:				
Total	0.141***	0.139**	0.156**	0.197
	[0.038]	[0.060]	[0.070]	[0.140]
Observations	606	606	606	606
Detected	0.064**	0.013	0.030	−0.085
	[0.027]	[0.044]	[0.052]	[0.092]
Observations	606	606	606	606
Panel B. Twins see each other a few times a year				
	(1) OLS	(2) All twins	(3) Fraternal twins	(4) Identical twins
Drug abuse	0.062*	0.046	−0.015	0.268**
	[0.038]	[0.055]	[0.066]	[0.110]
Observations	432	432	320	112
Drug dependence	0.088**	0.103**	0.105	0.222**
	[0.035]	[0.051]	[0.065]	[0.092]
Observations	432	432	320	112
Conduct disorder:				
APA definition	0.131**	0.118**	0.080	0.121
	[0.038]	[0.054]	[0.065]	[0.103]
Observations	432	432	320	112
Score	0.862**	0.650**	0.696	0.188
	[0.245]	[0.344]	[0.427]	[0.672]
Observations	432	432	320	112
Crime:				
Total	0.165**	0.103	0.155	0.333
	[0.067]	[0.120]	[0.141]	[0.464]
Observations	184	184	142	42
Detected	0.089**	−0.014	−0.016	0.243
	[0.036]	[0.080]	[0.087]	[0.323]
Observations	184	184	142	42

Note: Standard errors shown in brackets;* indicates statistically significant at 10% level, ** at 5% and *** at 1%

Despite the strong reduction of the sample we observe that the estimated effects remain quite similar to the findings in Table 2.4, especially when using all twins (columns (1) and (2)). The results in Panel A for fraternal and identical twins (columns (3) and (4)) are also very similar to the estimates in Table 2.4. In Panel B the sample sizes become very small for columns (3) and (4) but nearly all estimates suggest that child maltreatment increases illegal or problematic behavior. Hence, for a sample of twins for which we expect that spillover effects will be less likely the estimated effects are quite similar to our main estimation results. This suggests that spillover effects might not be a major concern for our estimates.

2.7 Conclusion

This chapter investigates the long-term effects of childhood maltreatment on illegal or problematic behavior using a sample of Australian twins. Previous studies have found a strong association between child maltreatment and illegal behavior which supports the so-called cycle of violence hypothesis. Here, we investigated whether the link between child maltreatment and illegal or problematic behavior also exists with pairs of twins and within pairs of identical twins.

We find consistent evidence that childhood maltreatment has a large effect on illegal and problematic behavior both between and within families. The estimated effects imply an increase of illegal or problematic behavior of fifty to one hundred percent relative to the baseline levels of individuals that have not been maltreated. This is consistent with the findings by Currie and Tekin (2012) who report a doubling of criminal behavior due to childhood maltreatment. Physical and sexual maltreatment have a large effect on all types of illegal or problematic behavior. Sexual maltreatment by an outsider increases externalizing behavior whereas sexual maltreatment by a family member induces internalizing behavior. For males we find that physical and sexual maltreatment increase all

types of illegal behavior. For females we find that both sexual maltreatment and physical maltreatment increase drug abuse (dependence) and conduct disorder.

In sum, our estimation results, based on OLS-models with a large set of controls and twin fixed effect models indicate a strong relationship between child maltreatment and illegal or problematic behavior. We acknowledge that our approach has potential limitations because the variation in child maltreatment is not derived from a randomized setting. In addition, the asymmetry in victimization within pairs of twins –'why just me?" – might generate additional resentfulness enhancing the 'cycle of violence'. However, our approach, which extends the control strategies applied in previous studies towards samples of twins and identical twins, might be considered as the most complete control strategy that is feasible. Therefore, we believe that our results should be interpreted as further evidence for the so-called 'cycle of violence' hypothesis. These results imply that child maltreatment not only has large private costs but also large social costs, which further legitimizes substantial governmental spending on prevention of child maltreatment and treatment of victims of child maltreatment. As we showed in the chapter, child maltreatment is more likely within families who have altogether more problems. Therefore, assisting children and parents from such families could be a good initial preventive measure. More information, centers, and a safe environment where victims of child maltreatment would feel comfortable to report about the event are another measure which could be helpful.

Chapter 3

Does victimization really increase offending[1]

3.1 Introduction

Many studies unravel a strong relationship between past victimization and current offending behavior (Wolfgang & Ferracuti (1967), Ousey *et al.* (2011), Lauritsen & Laub (2007), Smith & Ecob (2007)). Lauritsen and Laub (2007) even point that "The relationship has been found across time, place, and for various subgroups. It is significant regardless of the type of data used or the type of offending [...] or victimization [...] under consideration." A number of criminological theories explain the mechanism driving this relationship. However, selection into victimization is not random and studies which do not tackle the nonrandom selection problem should be viewed with caution. Given the absence of experimental data, an alternative option is to use an instrument as a source of exogenous variation. The difficulty with finding such an instrument is that most personal characteristics that affect the victimization status would also have an independent effect on the offending. Family background characteristics, peer influence and characteristics of the neighborhoods are likely to influence the criminal involvement. With the lack of a good instrument, an alternative would be to use longitudinal data (Ousey *et al.* (2011), Jennings *et al.* (2010), Schreck *et al.* (2006)). However, most existing panel studies do not

[1]This chapter is based on Misheva (2014). The paper greatly benefited from suggestions by Dinand Webbink, and by participants in the TI PhD seminars, attendees of the Warwick doctoral days (2015), the 15th EBES conference (2015) in Lisbon, and of the Workshop on Applied Microeconomics and Microeconometrics in Alicante (2015).

apply individual fixed effects (Jennings *et al.* (2010) use trajectory methodology; Smith & Ecob (2007) use latent growth mixture models), and even if individual fixed effects would be implemented, they cannot account for time-invariant characteristics, thus not necessarily ameliorating the endogeneity problem.

From a public policy perspective, crime is a very important issue because it involves large social costs. The costs could be both direct, such as expenses (medical costs and property damage), reduced productivity, costs of prevention and detection, or nonmonetary, such as pain and suffering, stress, mental discomfort and disorders (post-traumatic stress disorders, anxiety, fear), and overall a loss of quality of life. In the Netherlands, which is the country of focus in this paper, around 25% of citizens, on average, report to have been victims of crime, making the Netherlands the fourth country after Australia, New Zealand, and the UK in the crime victims table ("Netherlands Crime Stats", Nation Master; CBS 2009). [2]If only a fraction of this population would seek retaliation and offend, the costs to society could be enormous. In particular, the costs of crime in the Netherlands are already amounting to billions of euros per year (in 2005, they were around 20.2 billion euros, or 4.1% of GDP, see SEO (2007)).

Victimization may have numerous negative short and long-term consequences but in this paper we focus on one in particular – whether experiencing victimization leads to offending. Though other outcomes from victimization are important and deserve attention, we focus on offending because it is particularly costly and because of its potential to propel more victimization and offending in the long run (in a way, to create a snowball effect). Victimization as a traumatic experience (one can think of it as a negative treatment) could propel anger and urge the individual to retaliate and thus – offend. In criminology there are a few theories that explain what drives certain individuals and subgroups to retaliate and offend after becoming a target of either violent or a property crime, with adolescents being the most vulnerable (Jensen & Brownfield (1986), Hindelang *et al.* (1978), Sampson

[2]However, we need to keep in mind that this number could reflect willingness to report crime rather than a greater risk, as perception of safety in the Netherlands is among the highest in the world, and more importantly, the majority of these crimes are property-related or vandalism, violent crimes are much rarer.

& Lauritsen (1990), Ousey *et al.* (2011), Schreck *et al.* (2008), Agnew (1985), Agnew (2001)). Yet, none of the theories gives a complete explanation, nor do they agree on the magnitude of the effect. The difference in the findings could be due to the presence of confounding factors. It is likely that unmeasured and unobserved characteristics could affect the outcome. Establishing causality with offending behavior is especially difficult because randomized experiments are hardly applcable. We extend the findings from the current studies in the literature by using rigorous econometric approaches which are an improvement over the correlational analyses.

We distinguish between short- and long-run effect of victimization. If the offending took place shortly after the victimization, it is more likely to be impulsive, whereas the long-run effect might be associated with more careful planning. The hypothesis that we test in this paper does not necessarily imply rational and careful weighing of the costs and gains from offending (though, of course, that is still a possibility) because often retaliation could be spontaneous and unplanned. We use a representative sample from the Netherlands, the so-called Netherlands Survey of Criminality and Law Enforcement. The data set contains detailed information on current and past offending and victimization history, as well as rich personal background information. The challenge in an empirical study of offending is to find valid exogenous variation. Reducing and gauging the endogeneity of victimization is the main challenge we face in this paper. We tackle it in a few ways. We start with Ordinary Least Squares (OLS) with a large set of controls. Second, we focus on different samples, for which the victimization is (arguably) to a lower degree correlated with unobserved individual characteristics. Third, we use a strategy proposed by Altonji *et al.* (2005), in which they use selection on observables to obtain bounds for selection on unobservable factors. These methods rely on different identifying assumptions and using all of them will give us an idea about the relevance of selection bias. Of course, to be able to claim causality we would need to have clean exogenous variation, an experiment, or a good instrument, which is difficult when analyzing the effect of victimization. However, while neither of the approaches is superior in that it provides a clean control strategy,

using both of them allows us to gain insight into the significance and magnitude of the victimization-offending relationship.

Our findings can be summarized as follows. When we use OLS and include all the controls, we find that victimization (both in the previous year and the more distant past) increases the probability to offend in the current year with up to 1/3 compared to the baseline. These results are mostly driven by the male population. However, when we use Altonji's approach, we find a positive and significant upper bound but the lower bound is not significant, and is even negative.

This chapter is organized in the following way. In Section 3.2 we discuss some theories about the link between victimization and offending as well as some relevant studies. In Section 3.3 we describe our data and display some summary statistics. Section 3.4 explains the different estimation strategies we employ and their respective advantages and limitations. Section 3.5 presents results from our different estimation techniques, and Section 3.6 concludes.

3.2 Literature Review

3.2.1 Criminological theories about the link between victimization and offending

The theories in criminology and sociology that explain the link between victimization and offending can be summarized by two perspectives (Ousey *et al.* (2011)). The first one – the so-called *dynamic perspective* – suggests that the relationship between the victimization and offending goes in both directions. In general, such theories postulate that victimization changes individuals in a way that increases their risk for subsequent offending (Ousey *et al.* (2011)). One of the main theories that belongs to this stream of the literature is the subculture of violence. Proposed originally by Wolfgang & Ferracuti (1967), this theory suggests a victim-offender homogeneity and supports the idea that among certain

subcultures the use of violence is encouraged. Thus, people from such groups respond to violence with retaliation.

Agnew's general strain theory (1985, 2001, 2002) suggests that experiencing violence creates negative emotions (negative strain), most prominently anger, which leads to a desire for retaliation. But if victimization creates negative affect that later leads to offending, a negative strain is likely to be imposed on another victim, who will, in turn, be motivated to respond with offending as well (Ousey *et al.* (2011)). Not all strain would lead to crime. According to Agnew, certain conditions need to be met. These include acts of victimization that are seen as unjust, high in magnitude, associated with low social control, and creating an incentive to engage in criminal coping. Other possible responses from victimization are escapist (e.g. drug use) and instrumental (e.g., property offenses). Coping via illegal behavior and violence might be especially true for adolescents because of their limited legitimate coping mechanism and peer influence (Agnew (2001), Seepersad (2010)).

A strand of the literature focuses on child victimization and later delinquent outcomes. Studies explain why children as victims of violence later copy that behavior (so-called "cycle of violence" by Widom, 1989) by suggesting that children that are victims of abuse become apprehensive and fearful and respond with violence to stimuli from the environment (even when these stimuli are not negative or threatening). There are studies in economics who also show that child maltreatment leads to later criminal involvement (Currie & Tekin (2012), see also Chapter 2).

The second stream of theories is united around the so-called *population heterogeneity* perspective. It contends that there is a weak association between victimization and offending, and they correlated with unobserved individual characteristics. One such characteristic is the low self-control (Schreck (1999), Chen (2009)). Low self-control is rather stable throughout one's lifetime and is associated with lack for future orientation, hyperbolic discounting and tolerance for delinquent peers, which implies that individuals with low self-control would be exposed to a higher degree to deviant peers (Chen (2009)).

However, even though low self-control and other unobserved time-persistent characteristics might be related with both victimization and offending, criminology studies do not agree that accounting for population heterogeneity will completely explain the link between the two, but would most likely just give a better understanding about the dynamics between the two processes (Ousey *et al.* (2011)).

3.2.2 Economic perspective

Some of the criminological theories could be applied in an economic context. If offenders are maximizing their utilities, then given the risk of being caught and punished, they would choose attractive targets that would yield highest expected utility. Such could be people from higher social economic status, or people who are vulnerable because they live in dangerous neighborhoods, cannot afford proper guidance and are exposed to offenders. Offenders could be especially attractive targets because they are less likely to report a crime to the police when they are victimized in order not to be exposed for previous crimes.

The rational choice model, originally proposed by Becker (1968), explains rational offending behavior but some claim that it does not always fair well in reality (Entorf (2014)). Irrational behavior, such as retaliation, is not taken into account by it. Nor does it consider that a victim and an offender could be the same person. Victims might be spurred by anger and decide to take the law in their own hands if, for instance, their trust in the criminal and justice system is low. The threat of retaliation could deter future crimes, especially among certain subcultures. Retaliation, however, does not need to be directed to the perpetrator of a crime but could as well be randomly targeted at a member of a group with the purpose of deterring more offenses in the future (Jacobs & Wright (2010), Agnew (2001)). For vulnerable individuals sometimes it might be beneficial to join a sub-group in order to feel protected. One study (see Sobel & Osaba (2009)) argues that youth might join gangs due to government's failure to protect them.

Other factors could also be important for the decision-making process. Such factors, as pointed by bounded rationality theory, could be emotions (especially anger and desire for retaliate and restore justice), patience, time preference, hyperbolic discounting, and others. For instance, anger and the desire to retaliate could stem from norms for honor and respect (also pointed out by many of the criminological theories above), and it might not be directed to the perpetrator, but also to non-involved potential victims, and even at a later point in time (Jacobs & Wright (2010)).

3.2.3 Previous studies about the link between victimization and offending

One of the first studies on this topic was conducted by Singer (1981, 1986) who used data from the Wolfgang's Philadelphia birth cohort. He found that victims were between 1.7 and 2.8 times more likely to report offending behavior than non-victims. This risk is increased to 17 times when he focused solely on gang members. He explains his findings with the influence of peers, the lack of social controls (such as family members, teachers, authority figures), and social learning (i.e., individuals legitimatize delinquent behavior by the notion "everyone is doing it").

Agnew (1985) formulated the hypothesis about the effect of strain on delinquency. He used the Youth in Transition Survey from 1966, conducted among tenth-grade boys and found that strain has a (direct and indirect) effect on both property offenses and violent offenses. Cullen et al. (2008) confirm the validity of the general strain theory proposed by Agnew, using data from Virginia. They find that being a victim of bullying is positively associated with delinquency and substance use. The effect is stable across genders and is stronger for students with weak social bonds. Hay and Evans (2006) find that victimization is associated with later delinquency, even after controlling for previous offending. Piquero and Sealock (2000) conducted a study among incarcerated youths on the effects of both anger and depression on violent and property crimes. They find that depression fails to predict both crimes, whereas anger is associated with violence but not

with property crime.

A strand of the literature studies the long-term effect of being a victim of abuse (of physical, sexual, emotional) during childhood on later deviant behavior. Widom (1989) is among the first researchers to propose the cycle of violence hypothesis. She matched a sample of substantiated cases of child abuse with samples from the general population to find that children who were victims of child abuse have, on average, around 50% higher probability to display delinquent behavior in comparison to their comparable but not-abused peers. They also have higher probability to be arrested and go to jail. More recently, Currie and Tekin (2012) used data from the US and sibling fixed-effects to find that neglect and sexual abuse, and to a lower degree the physical abuse, double the risk to commit a number of crimes. In Chapter 2, we also showed that child maltreatment increases the probability to display problematic behavior between fifty and one hundred percent.

However, there are studies which argue that the link between the two might be driven by unobseved confounders. Ousey et al. (2010) use longitudinal data on middle and high-school students to find that in models controlling for time-stable individual characteristics there is a negative relationship between victimization and offending.

The reviewed studies look at the relationship between past victimization and offending from a criminological and sociological perspective. Besides Currie and Tekin (2012) who focus on a specific case of victimization (child maltreatment) and our own study in Chapter 2, which uses twins data, to the best of our knowledge there are no other studies in economics that investigate this problem. This analysis contributes to the literature by testing these criminological theories using a representative sample among the Dutch population, distinguishing between short-run and long-run effect of past victimization.

3.3 Data

3.3.1 Data description

We use data from the Netherlands Survey on Criminality and Law Enforcement, conducted in 1996 on a random sample of the Dutch population above the age of 15. Altogether the sample contains 2951 observations.[3] The survey includes questions about the age, marital status, education, employment, and family history. Interviewees were asked retrospectively about their victimization and offense experiences. Those who report any victimization/offense acts, were further inquired details about each of these experiences. It is important to note that for each respondent we know not only the latest victimization/offense act (which did not necessarily occur at the year of the interview), but we also know the first victimization and offending the respondent reported, and which could have happened several years in the past. A detailed description of the data can be found in Wittebrood & ter Voert (2007) and Wittebrood & Nieuwbeerta (1999).

Since the data are retrospective, a number of problems could threaten the precision of the collected information. Various measures were taken to reduce potential issues. During the interviews, respondents made use of the so-called "event history calendars", in which they marked important events (such as marriages, births of children, year of first employment, etc.) and then they were asked to place the victimization and offending history making use of these calendars. This method has been implemented to reduce the measurement error due to memory decay of the respondents. To reduce the threat of a measurement error due to misreporting (sometimes respondents might be tempted to choose a socially desirable answer), respondents answered the offending-related questions directly to the computer (and not to the interviewer), and it was not possible to skip questions. Of course, the social desirability with regards to the offending questions could still be present, despite the manner in which the interview was conducted. And whereas

[3]The data collection followed a few steps. First, municipalities were selected that were representative of Dutch municipalities with regards to region and degree of urbanization. In the second step, using the national mail delivery register, addresses were picked. The final step was to select the respondent. In each of the addresses, the person, above 15, whose birthday was coming up next was the one interviewed.

we cannot calculate how large this potential threat would be, if someone reported a false negative (committed a crime but said they did not), this would lead to a downward bias in our estimates.

Our main outcome variable is a dichotomous one and accounts for the current offending. It takes a value of 1 if the respondent reported committing any one of a number of crimes in 1996 – the year the interview took place. Our main "treatment" variable takes a value of 1 when the respondent reported to be a victim of a crime. We investigate separately the short-run and the long-run effect of victimization. For the short-run, we construct a dichotomous variable capturing victimization in 1995, the year before the interview.[4] When we look at the long-run effect, we construct a variable for victimization that took place any time before 1995[5]

We include a large set of control variables in our analysis that are important correlates of offending behavior. These are the age of the respondent, indicator variables for being married, being a male and having children (and their number), being born in the Netherlands, being unemployed, having only primary education or a college one, being religious, having bad relationship with either of the parents and with the (former) teachers. We further have a variable accounting for the criminal activity in the neighborhood of the respondent. This is a self-reported variable and could reflect the respondent's perception of safety, not necessarily the overall safety of the neighborhood. A variable measures complications at birth and it takes a value of 1 when the respondent has experienced difficulties at his/her own birth. We include a variable for behavior differences with adults, and another which refers to any such differences that led to consultations with a psychologist or a psychiatrist. The latter is important because problematic behavior often starts in early

[4]We have victimization information for 1996 as well, but we want to make sure the "treatment" preceded the outcome.

[5] The offending is defined as 1 when the respondent replied he/she committed tax, social security or insurance fraud, committed theft from work, from a house or a car, stole money, inflicted an injury with a weapon, assaulted someone, threatened someone, committed hit-and-run, stole a bicycle, was involved in fencing, vandalized private or public property, or switched price tags.

The victimization is defined as 1 when the respondent stated that he/she experienced any of the following: assault, threat, sexual offense, burglary, a bike or a car theft, pickpocketing, or some other kind of theft, vandalism, hit-and-run, telephone harassment, or some other type of a crime, to be specified by the victim him/herself.

age and is persistent over one's life.

3.3.2 Descriptive statistics

Table 3.1 shows the descriptive statistics and the offending prevalence in our sample. The first column shows the whole sample of respondents, the second column displays those who report any victimization in 1995 and the last two those who did not and the difference in the means, respectively. Around 30% of the whole population reports being a victim in 1996, and about 31% being a victim in 1995. As already mentioned, on average, around 25% of the Dutch population reports to be a victim of a crime (CBS 2009, Nationmaster Netherlands statistics). The incidence rate in our sample then is very close to the one in these reports. Around 12% reported to have committed an offense in 1996, and 14% in 1995, the majority of which are property-related crimes. Official statistics find that among adolescents, around 40% have committed some offense (CBS, Jeugdcriminaliteit in de periode 1995-2010), and on average, around 2% of the Dutch population are suspects of a crime (SCP, Criminaliteit, hoofdstuk 8). At the same time, only about 33% of crimes are reported to the police, and from those reported, not all are recorded (SCP, Criminaliteit), the offending rates in our sample then do not look extreme. We observe large differences in offending both in 1996 and in 1995 between columns (2) and (3). Those who report being a victim in 1995 are around 8% more likely to offend in 1996, and they were around 12% more likely to offend in 1995. Though not shown in the table, we failed to find substantial difference in current offending between those who were victims at any point in the past and those who were not. These associations lead us to expect a stronger short-run effect of victimization and a weaker one in the long run.

Victims differ from non-victims in many of their observable characteristics as well. They are substantially younger, less frequently married, have fewer children, more often unemployed and less educated, live in neighborhoods with higher criminal rate, more often have a bad relationship with their father and (former) teachers, are less frequently

religious, are more likely report having risky lifestyles, had behavioral differences with adults in their childhood, and consulted a psychiatrist.

Table 3.1 Descriptive statistics and difference in means

	All respondents	Victim in 1995	Not a victim in 1995	T-test column(2)-column(3)
	(1)	(2)	(3)	(4)
Victim in 1996	0.30			
	[0.46]			
Victim in 1995	0.31			
	[0.46]			
Offend in 1996	0.12	0.17	0.09	0.08**
	[0.46]	[0.38]	[0.29]	(0.01)
Offend in 1995	0.14	0.22	0.10	0.12***
	[0.46]	[0.42]	[0.30]	(0.01)
Age	36.9	31.73	39.34	−7.62***
	[17.6]	[15.62]	[18.06]	(0.68)
Male	0.45	0.44	0.48	−0.01
	[0.50]	[0.50]	[0.50]	(0.02)
Married	0.44	0.36	0.48	−0.12***
	[0.50]	[0.48]	[0.50]	(0.02)
Children (1 if yes, 0 if none)	0.47	0.38	0.51	−0.13***
	[0.50]	[0.48]	[0.50]	(0.02)
Number of children	1.09	0.81	1.21	−0.40***
	[1.42]	[1.25]	[1.48]	(0.06)
Born in the NL	0.95	0.94	0.95	−0.01
	[0.22]	[0.23]	[0.22]	(0.009)
Unemployed	0.03	0.04	0.03	0.01**
	[0.19]	[0.21]	[0.17]	(0.007)
College education	0.18	0.19	0.17	0.02
	[0.38]	[0.39]	[0.37]	(0.02)
Primary education	0.21	0.19	0.23	−0.03**
	[0.41]	[0.39]	[0.42]	(0.02)
High crime rate in neighborhood	0.14	0.23	0.10	0.13***
	[0.35]	[0.42]	[0.31]	(0.01)
Bad relat. with father	0.03	0.05	0.02	0.03**
	[0.18]	[0.21]	[0.15]	(0.007)
Bad relat. with mother	0.02	0.03	0.02	0.01
	[0.14]	[0.16]	[0.13]	(0.006)
Bad relat. with teachers	0.01	0.014	0.007	0.007*
	[0.09]	[0.12]	[0.09]	(0.004)
Religious	0.50	0.45	0.52	−0.07**
	[0.50]	[0.50]	[0.50]	(0.02)
Risky lifestyle	0.57	0.61	0.54	0.06**
	[0.50]	[0.49]	[0.50]	(0.02)
Complications at birth	0.08	0.09	0.07	0.01
	[0.27]	[0.28]	[0.26]	(0.01)
Behav. differences with adults	0.08	0.10	0.07	0.03**
	[0.27]	[0.31]	[0.25]	(0.01)
Behavior problems: psychologist	0.02	0.03	0.01	0.016**
	[0.013]	[0.16]	[0.11]	(0.006)
Observations	2951	925	2026	

Note: standard deviations are given in brackets []; standard errors are in parentheses ();
* indicates statistically significant at the 10% level, **- at the 5% level, and ***- at the 1% level

3.4 Empirical Strategies

3.4.1 OLS (with different samples)

We want to distinguish between the short- and long-run effects of victimization since each type can initiate a different behavioral reaction (for instance, a spontanenous one in the short run versus a better rationalized one in the long run). In addition, victimization before 1995 could affect the probability to be victimized in 1995, which could make the effect of victimization before 1995 misleadingly low. Therefore, we estimate two separate equations. We start with a simple OLS model with an outcome offending in 1996 ($O_{i,1996}$), explained first only by victimization that took place in 1995 (i.e. short-run victimization), then we add control variables and estimate an equation of the following form:

$$O_{i,1996} = \alpha_0 + \alpha_1 V_{i,1995} + \alpha_2 X_i + e_i \tag{3.1}$$

where X_i is a vector of personal characteristics, displayed in Table 1, and e_i is an error term. We further estimate (1) separately for males and females.

For the long-run effect of victimization on offending, the equation we estimate is:

$$O_{i,1996} = \beta_0 + \beta_1 V_{i,<1995} + \beta_2 X_i + u_i \tag{3.2}$$

where $V_{i,<1995}$ stands for victimization before 1995, X_i is the same vector of characteristics and u_i is an error term. The hypothesis we will test in both equations is whether α_1 and β_1 are positive and significant correlates of the outcome. However, a major concern with a naive estimation such as (3.1) and (3.2) is that there is unobserved individual heterogeneity, correlated with being a victim and with the offending behavior, which will bias the regression results. If it is more likely that the unobserved heterogeneity increases the probability to become a victim and subsequently to offend, the endogenous variable will capture this effect as well, leading to an upwardly biased coefficient (if there is a negative relationship, the coefficient will be downward biased). To reduce the omitted

variables bias, we include a rich set of control variables, detailed in Section 3.3. However, such a control strategy might not eliminate the endogeneity problem. We further try to reduce the selection bias by looking at specific subsamples. We try to eliminate factors that have been found important in both the victimization and offending. For instance, a risky lifestyle could be related to both a higher probability to be a victim and can suggest criminal tolerance. Thus, by excluding people with risky lifestyle, we could remove this particular source of selection bias. We expect among each of the subsamples we look at, that the victimization would be to a lower extent correlated with observed and unobserved characteristics. First, we focus on those who report not to have a risky lifestyle. One could argue that people with risky lifestyle are more vulnerable targets and the risky behavior might be indicative of delinquent association. Second, we focus on individuals who did not know the perpetrator of the crime. Knowing the perpetrator could indicate that the respondent already had links to criminal individuals. Third, we look at those who became a victim of a crime but did not report it to the police. Our hypothesis here is that the crime would not be reported either because it was not serious enough, the respondent does not have faith in the police, or because the victim him/herself has offense history and does not want to be exposed to the police. However, we need to acknowledge that even when we focus on these samples, the selection bias might not be completely removed.

3.4.2 Altonji et al. (2005)'s strategy

Given the potential shortcomings of the OLS estimation, we implement a second approach. It follows the strategy outlined by Altonji et al. (2005), and later applied to other papers (Dujardin & Goffette-Nagot (2009), Ward & Williams (2014), for instance). In their study of the effectiveness of Catholic schools on test scores and educational attainment, Altonji et al. (2005) develop a new estimation approach following the idea that the amount of selection on the observed explanatory variables provides a guide for the selection on unobservables.

They start by employing a bivariate probit model in a two-step procedure, where due to the lack of a proper instrument, the correlation coefficient ρ between the error terms of the two equations is treated as unidentified. In our case, the bivariate probit model for the long-run effect of victimization (and an identical one for the short-run effect where we have the variable $V_{i,1995}$ instead) is a system such as (3.3), (3.4) and (3.5):

$$O_{i,1996} = 1(X'\gamma + \alpha V_{i,<1995} + \eta) \tag{3.3}$$

$$V_{i,<1995} = 1(X'\beta + \nu) \tag{3.4}$$

$$\begin{bmatrix} \eta \\ \nu \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right) \tag{3.5}$$

In the first step, Altonji et al. (2005) estimate equations such as (3.3) and (3.4) simultaneously while varying the value of the correlation coefficient ρ and at the same time – checking the sensitivity of the estimated effect of the endogenous variable to the degree of correlation in unobservables. In this way, ρ is treated as if not identified and the goal is to find the value of ρ at which the endogenous variable no longer has an effect on the outcome (in other words, the threshold for which the degree of selection on unobservables for which the effect of victimization on offending disappears). In our case, if unobservable factors (such as risky lifestyle, having delinquent peers, etc.) that raise the probability to become a victim also increase the probability to offend, then we will expect a positive correlation and use positive increments of ρ. Therefore, following Altonji et al. (2005) (this approach was originally suggested by Rosenbaum & Rubin (1983)), we will estimate a bivariate probit with $\rho = 0$ (when we are in the univariate case), and values of 0.05, 0.10, 0.15, and so forth.[6]

The second step in Altonji et al.'s methodology is to find bounds for the correlation

[6] Though it is possible that the effect of victimization on the outcome is not monotonous i.e. for some respondents previous victimization decreases the probability to offend, given what we observe in our data so far and the positive relationship found in previous studies, we expect a positive correlation coefficient ρ.

coefficient and the corresponding endogenous variable. A lower bound is obtained from imposing the assumption that selection on unobservables is equal to selection on observables, and an upper bound is obtained when assuming that ρ is zero, i.e. the victimization variable (either short- or long-run one) is exogenous and the two equations can be estimated by a univariate probit. If we ignore the endogeneity of the victimization, then the estimated coefficient will also capture the effect of the unobserved factors, producing an upwardly biased coefficient. Altonji et al. (2005) show that in the bivariate probit case the selection on observables equating selection on unobservables is equivalent to:

$$\hat{\rho} = \frac{Cov(X'\beta,\, X'\gamma)}{Var(X'\gamma)} \tag{3.6}$$

It is more likely, however, that there will be a higher selection on observables than on unobservables because the purpose of all surveys is to collect useful information for the respective problem in case; and given the high number and relevance of the covariates we include, it is very plausible that the selection on the observables is higher than the selection on the unobservables.[7] Therefore, we would expect the estimate of the endogenous variable based on the assumption that there is equal selection on observable and unobservable factors to be downward biased. The bounds for ρ that effectively bound the coefficient of interest, could be given by the following inequality:

$$0 \leq \rho \leq \frac{Cov(X'\beta,\, X'\gamma)}{Var(X'\gamma)} \tag{3.7}$$

Altonji's approach relies on three identifying assumptions. First, the elements of X are chosen at random from the full set of observable and unobservable factors that determine the outcome. Second, the number of elements in X are large and none dominates the distribution of victimization or the offending. The third assumption is that the selection on observables is equal to the selection on unobservables. While the authors point that these are indeed strong assumptions that are unlikely to hold exactly, they argue that they

[7] The approach of Altonji et al. (2005) relies on the assumption of higher selection on observables than on unobservables and without it, a lower bound cannot be estimated.

are not any stronger than the standard OLS assumptions.[8]

Before they theoretically and empirically apply this method, the authors explore some nice features of their data. First of all, they have a broad set of explanatory variables, which cover most of the important factors in the literature. In our case, we also have a broad set of characteristics – familial, personal, and neighborhood ones – that are included in the majority of offending and victimization studies. Furthermore, they take advantage of the timing in their data and the decision to attend Catholic school predates the outcome. In our case, the endogenous variable took place before the outcome. Finally, they are able to isolate a relatively homogenous sample from their population. In particular, when studying the probability to graduate from high school (or college), they can focus on a sample of Catholic eight-graders who decide whether to attend a Catholic or a public high school. Altonji et al. rely on the fact that almost none of the public school eight-graders continue in Catholic high school, and for these Catholic school eight-graders, there are smaller differences in the observable characteristics. We need to stress though that even for this homogenous sample, there is still a selection bias albeit a lower one. In our case, a homogenous sample is not so easy to isolate because the probability to become a victim could vary in each period (as opposed to a one-time decision – whether to attend Catholic high school or a public one).[9]

[8] For instance, if V* and O* are the latent variables for victimization and offending, the assumption of equal selection on observables and unobservables implies that the part of V* that is related to the unobservables and the part of V* related to the observables have the same relationship with O*. Whereas the OLS assumption would be that the part of V* that is related to the unobservables has no relationship with O*.

[9] In order to isolate a homogenous sample, we focused on people who report to have risky life, to know the perpetrator of the crime, or define high opportunity to offend because arguably they would have higher predisposition to offend. We also looked at samples with lower likelihood to commit crimes (such as reporting first-time victimization, not having risky lifestyle, reporting no victimization in 1995), but these sample were again heterogeneous in their likelihood to engage in delinquent behavior and to be victimized. We also focused on a younger sample (15-20 years old) but the findings were very similar to those for the main sample.

3.5 Empirical results for the effect of past victimization on current offending

3.5.1 Short-run effect of past victimization on current offending: OLS results

We start with presenting results from an OLS estimation of the short-run effect of past victimization on current offending. Table 3.2 displays the first set of results. Our independent variable of interest is victimization in 1995. In column (1) of Table 3.2, we include no control variables and we see victimization in 1995 is positively and significantly associated with current offending. The coefficient is still positive and significant when we add all the controls (see Column (2)), though it reduces in magnitude to about 3.7 percentage points. Compared to the baseline offending rate of 12 percentage points for the whole population, this coefficient indicates a substantial effect. From the other covariates, being a young male, who is not married, was born in the Netherlands, does not have a college degree, lives in a neighborhood with high crime rate and is not religious are significantly associated with the current offending behavior.

Columns (3) and (4) present the results by gender. The coefficient of victimization in 1995 is positive for both groups but it is statistically significant (at the 1% level) only for males. If a male reported to be a victim of a crime in 1995, this is associated with around 7 percentage points higher probability that he offends in 1996. It seems that males were driving the result we found in the first two columns.

Table 3.2 Effect of victimization in 1995 on current ofending

	Offending in 1996	Offending in 1996	Offending in 1996, males	Offending in 1996, females
	(1)	(2)	(3)	(4)
Victim in 1995	0.078***	0.037***	0.072***	0.009
	[0.013]	[0.013]	[0.022]	[0.014]
Age		−0.003***	−0.005***	−0.002***
		[0.000]	[0.001]	[0.000]
Male		0.071***		
		[0.012]		
Married		−0.063***	−0.033	−0.072***
		[0.015]	[0.028]	[0.016]
Any children		−0.008	0.018	−0.024
		[0.022]	[0.042]	[0.023]
Number of children		0.002	−0.008	0.007
		[0.007]	[0.014]	[0.007]
Born in NL		0.054**	0.111**	0.021
		[0.025]	[0.048]	[0.028]
Unemployed		0.003	0.044	−0.030
		[0.031]	[0.049]	[0.040]
College degree		−0.029*	−0.042	−0.010
		[0.015]	[0.026]	[0.018]
Primary educ.		0.016	0.036	−0.006
		[0.014]	[0.025]	[0.017]
High crime in neighborhood		0.051***	0.011	0.076***
		[0.016]	[0.030]	[0.018]
Bad rel. father		0.029	0.032	0.044
		[0.033]	[0.060]	[0.038]
Bad rel. mother		0.048	0.009	0.066
		[0.041]	[0.077]	[0.045]
Bad rel.teachers		0.063	0.048	0.076
		[0.060]	[0.088]	[0.083]
Religious		−0.030**	−0.043**	−0.021
		[0.012]	[0.020]	[0.014]
Problematic birth		−0.019	−0.058	0.010
		[0.021]	[0.036]	[0.025]
Beh.differ.adults		0.005	0.023	−0.015
		[0.024]	[0.035]	[0.032]
Consulted psych.		−0.000	0.012	−0.022
		[0.050]	[0.074]	[0.068]
Observations	2951	2951	1315	1636

Note: Standard erros in parantheses; * indicates statistically significant at 10%, ** at 5%, and *** at 1 %;

As a next step, we explore the short-run effect of victimization on current offending for different subsamples from our population, for which there is arguably a lower degree of selection bias. The results are presented in Table 3.3. In the first column we include individuals with no risky lifestyles since people with risky lifestyles are presumably more vulnerable targets. The victimization coefficient is not significant. In column (2) we explore the subsample of those who did not know the perpetrators. The victimization

coefficient is not statistically significant in this case either, and it even becomes negative. Column (3) contains those with no risky lifestyle and who did not know the perpetrators of the crimes committed against them. The victimization variable is still negative and not significant. Finally, the last column displays the sample of those who did not report the crime to the police, which could indicate a criminal past for which they do not want to be exposed. Again, we see no effect of past victimization on the current offending. The positive effect we saw in Table 3.2 disappears when we focus on samples for which there is arguably lower selection bias, possibly indicating that the positive effect was capturing unobserved heterogeneity (positively correlated with the offending variable). However, we need to acknowledge that our samples have also decreased as a result from performing this exercise.

Table 3.3 Effect of victimization in 1995 on current offending for different samples

	No risky lifestyle	Does not know perpetrator	No risky lifestyle and does not know perpetrator	Did not report crime to police
Victim in 1995	0.020	−0.014	−0.036	−0.020
	[0.016]	[0.032]	[0.046]	[0.025]
Age	−0.002***	−0.005***	−0.005***	−0.005***
	[0.000]	[0.001]	[0.002]	[0.001]
Male	0.043***	0.146***	0.112***	0.103***
	[0.015]	[0.029]	[0.047]	[0.024]
Married	−0.033*	−0.102***	−0.077	−0.102***
	[0.017]	[0.038]	[0.052]	[0.033]
Any children	−0.015	−0.022	0.032	−0.034
	[0.025]	[0.065]	[0.084]	[0.052]
Number of children	−0.003	0.000	0.000	0.006
	[0.008]	[0.025]	[0.030]	[0.018]
Born in NL	0.066**	0.074	0.173*	0.106*
	[0.029]	[0.065]	[0.088]	[0.049]
Unemployed	0.014	0.097	0.093	−0.014
	[0.040]	[0.073]	[0.118]	[0.059]
College degree	0.046**	−0.051	0.077	−0.049
	[0.021]	[0.038]	[0.060]	[0.031]
Primary educ.	−0.004	0.030	0.045	−0.028
	[0.017]	[0.038]	[0.056]	[0.030]
High crime in neighborhood	0.056***	0.032	0.086	0.081***
	[0.021]	[0.035]	[0.053]	[0.029]
Bad rel. father	−0.014	0.058	0.004	0.049
	[0.042]	[0.073]	[0.109]	[0.059]
Bad rel. mother	0.009	0.058	0.086	0.101
	[0.046]	[0.093]	[0.121]	[0.075]
Bad rel.teachers	0.003	0.185	0.000	0.149
	[0.085]	[0.119]	[0.168]	[0.105]
Religious	−0.004	−0.035	0.033	−0.014
	[0.015]	[0.030]	[0.044]	[0.025]
Problematic birth	−0.042	0.039	−0.122*	−0.011
	[0.027]	[0.050]	[0.068]	[0.044]
Beh.differ.adults	0.008	0.059	0.186**	0.022
	[0.033]	[0.059]	[0.091]	[0.046]
Consulted psych.	0.002	−0.084	−0.117	−0.013
	[0.060]	[0.107]	[0.145]	[0.087]
Observations	1283	644	249	1006

Note: Standard erros in parantheses; * statistically significant at 10%, ** at 5%, and *** at 1 %;

3.5.2 Using Altonji et al.'s approach for the short-run effect of victimization on offending

We continue with the second estimation technique we employ– a bivariate probit model,

treating the correlation coefficient between the error terms of the two equations as unidentified. In their first step, Altonji et al. vary the value of ρ to gauge at what level the degree of selection on unobserved factors no longer matters.

The results are given in Table 3.4. Each column presents the result of a different bivariate probit model that includes the whole set of control variables. Note that we include a large set of control variables, which importance has been stressed in the majority of offending studies. When we assume no correlation, i.e. when $\rho = 0$, the short-run effect of victimization is positive and statistically different from zero, with a marginal effect of 0.027. This is comparable in magnitude to the estimate from Column (2) in Table 3.2. We see as we move from the first column, the endogenous variable is no longer statistically significant, and becomes negative for $\rho = 0.15$. Adding only little selection bias makes the effect insignificant. Imposing a higher positive correlation of the error terms assumes that the unobserved characteristics that increase the risk for victimization also increase the probability to offend. In other words, the higher the sorting on unobservables, the lower is the effect of victimization.

In the last column, we impose the equal selection on observables and unobservables assumption. As discussed, the value of ρ for which the selection on observables is equal to the selection on unobservables provides a lower bound. In this case, the equal selection is at $\rho = 0.142$. The corresponding marginal effect of victimization is negative and not statistically significant. Having a negative effect of the endogenous variable could indicate that the hypothesis of equal selection on observables and unobservables is too strong in our case. We include a lot of personal characteristics that are important for the offending behavior. As a result, perhaps there is a lower selection on unobservables than on observables in our case. The upper bound is when $\rho = 0$, i.e. when we assume that there is no correlation between the two equations. When $\rho = 0$, the victimization in 1995 has a marginal effect of 0.027. The zero value is in the confidence interval for the short-run effect of victimization based on the two extreme hypotheses of equal and no selection. Given the positive effect of past victimization found in the literature and in our estimates so far,

such a negative effect if unlikely and probably due to the equal selection assumption.

Further, we followed the approach provided in Altonji et al. (2005) to calculate the relative amount of selection on unobservables relative to the selection on observables that is required to explain the whole effect of victimization on offending. We calculated this ratio to be 0.173. In other words, a low selection on the unobservables relative to selection on the observables is enough to explain the positive effect of victimization.

Table 3.4 Estimates of victimization in 1995 on offending with different values of ρ

	$\rho = 0$	$\rho = 0.05$	$\rho = 0.1$	$\rho = 0.15$	$\rho = 0.2$	$\rho = \hat{\rho} = 0.142$
Victimization in 1995	0.164^{**} [0.074]	0.082 [0.070]	0.001 [0.070]	-0.081 [0.069]	-0.162^{**} [0069]	-0.068 [0.070]
Marginal effects	**[0.027]**	[0.013]	[0.0001]	[-0.013]	[-0.027]	**[-0.011]**

Notes: Robust standard errors are given in brackets[];
** means statistically significant at the 10% level, ** at the 5% level and *** at the 1% level*

3.5.3 Long-run effect of past victimization on current offending: OLS results

In this section, we focus on the long-run effect of victimization on the current offending behavior. The results for the whole sample are presented in Table 3.5. In column (1) we include no control variables and we see past victimization is positive but not statistically significant. In column (2) we add control variables and the coefficient of past victimization increases in magnitude and becomes statistically significant. If someone reported to be a victim of a crime before 1995, this is associated with around 3 percentage points increase in the probability to offend in the current period. This coefficient is rather close to the coefficient of the short-run effect of victimization, presented in Table 3.2, and again, a substantial effect compared to the baseline offending reported by 12% of the population. Of course, given the high threat of omitted variables bias, we should take this coefficient with caution. In general, the same covariates as in Table 3.2 are correlated with the delinquent behavior. Being a young, unmarried male, without a college degree, who lives in a dangerous neighborhood and is not religious, are statistically significant associates of the current offending. Columns (3) and (4) again present the results by gender. The coefficient of past victimization is positive in both columns, and only for males it is significant at the

10% level, though it is not as distinctly pronounced as the short-run one.

As a next step, we explore the long-run effect of victimization on current offending for the different subsamples from our population. The results are given in Table 3.6. For all of the groups we find a small, positive coefficient, but only for the subsample of those who did not know the perpetrator of the crime, is the effect significant at the 10% level. The small positive effect from Table 3.5 disappears when we focus on these distinct samples. If this method credibly removed the selection bias, we still would not find evidence that a victimization act in the past increases the current probability of delinquent behavior. However, since our samples are smaller, we refrain from making strong conclusions.

Table 3.5 Effect of victimization before 1995 on current offending

	Offending in 1996	Offending in 1996	Offending in 1996, males	Offending in 1996, females
	(1)	(2)	(3)	(4)
Victim in 1995	0.015	0.032***	0.049*	0.013
	[0.015]	[0.015]	[0.025]	[0.018]
Age		−0.003***	−0.005***	−0.003***
		[0.000]	[0.001]	[0.000]
Male		0.071***		
		[0.012]		
Married		−0.067***	−0.039	−0.073***
		[0.015]	[0.028]	[0.016]
Any children		−0.010	0.016	−0.025
		[0.022]	[0.042]	[0.024]
Number of children		0.003	−0.006	0.007
		[0.007]	[0.014]	[0.007]
Born in NL		0.049**	0.107**	0.019
		[0.025]	[0.048]	[0.028]
Unemployed		0.004	0.040	−0.028
		[0.031]	[0.049]	[0.040]
College degree		−0.030*	−0.047*	−0.010
		[0.015]	[0.026]	[0.018]
Primary educ.		0.019	0.041*	−0.005
		[0.014]	[0.025]	[0.017]
High crime in neighborhood		0.056***	0.028	0.076***
		[0.016]	[0.029]	[0.018]
Bad rel. father		0.031	0.033	0.044
		[0.034]	[0.060]	[0.038]
Bad rel. mother		0.049	0.006	0.066
		[0.041]	[0.077]	[0.045]
Bad rel.teachers		0.063	0.046	0.075
		[0.060]	[0.089]	[0.083]
Religious		−0.030**	−0.043**	−0.020
		[0.012]	[0.020]	[0.014]
Problematic birth		−0.021	−0.058	0.009
		[0.021]	[0.036]	[0.025]
Beh.differ.adults		0.007	0.026	−0.015
		[0.024]	[0.035]	[0.032]
Consulted psych.		−0.001	0.015	−0.022
		[0.050]	[0.074]	[0.068]
Observations	2951	2951	1315	1636

Note: Standard erros in parantheses; * indicates statistically significant at 10%, ** at 5%, and *** at 1 %;

Table 3.6 Effect of victimization before 1995 on current offending for different samples

	No risky lifestyle	Does not know perpetrator	No risky lifestyle and does not know perpetrator	Did not report crime to police
Victim in 1995	0.023	0.007*	−0.041	0.049
	[0.018]	[0.038]	[0.059]	[0.025]
Age	−0.002***	−0.005***	−0.005***	−0.005***
	[0.000]	[0.001]	[0.002]	[0.001]
Male	0.043***	0.148***	0.116***	0.106***
	[0.015]	[0.029]	[0.052]	[0.024]
Married	−0.035**	−0.111***	0.030	−0.106***
	[0.017]	[0.038]	[0.084]	[0.033]
Any children	−0.017	−0.022	0.032	−0.035
	[0.025]	[0.064]	[0.084]	[0.052]
Number of children	−0.003	0.001	0.002	0.007
	[0.008]	[0.025]	[0.030]	[0.018]
Born in NL	0.062**	0.055	0.163*	0.093*
	[0.029]	[0.066]	[0.089]	[0.050]
Unemployed	0.015	0.101	0.088	−0.016
	[0.040]	[0.072]	[0.118]	[0.059]
College degree	0.045**	−0.057	0.073	−0.053*
	[0.021]	[0.038]	[0.060]	[0.031]
Primary educ.	−0.002	0.037	0.052	0.034
	[0.018]	[0.038]	[0.057]	[0.030]
High crime in neighborhood	0.059***	0.026	0.077	0.077***
	[0.020]	[0.035]	[0.053]	[0.029]
Bad rel. father	−0.015	0.050	−0.008	0.044
	[0.042]	[0.073]	[0.110]	[0.059]
Bad rel. mother	0.010	0.061	0.089	0.102
	[0.046]	[0.093]	[0.121]	[0.075]
Bad rel.teachers	0.007	0.171	−0.014	0.143
	[0.084]	[0.119]	[0.168]	[0.105]
Religious	−0.004	−0.034	0.039	−0.014
	[0.015]	[0.030]	[0.044]	[0.025]
Problematic birth	−0.043	0.039	−0.122*	−0.009
	[0.027]	[0.050]	[0.068]	[0.044]
Beh.differ.adults	0.008	0.062	0.185**	0.023
	[0.033]	[0.058]	[0.091]	[0.046]
Consulted psych.	0.005	−0.089	−0.112	−0.013
	[0.060]	[0.106]	[0.145]	[0.087]
Observations	1283	644	249	1006

Note: Standard erros in parantheses; * is statistically significant at 10%, ** at 5%, and *** at 1 %;

3.5.4 Using Altonji et al.'s approach for the long-run effect of victimization on offending

Finally, we repeat Altonji et al.'s approach for the long-run effect of victimization on

the current offending. The results are given in Table 3.7. As above, the first five columns show the effect of the endogenous variable as we vary the level of correlation between the observable and unobservable variables. When we assume no correlation, i.e. $\rho = 0$, the long-run effect of victimization is positive and statistically significant with a marginal effect equal to 0.038, which is very similar to the OLS coefficient in Column (2) of both Tables 3.2 and 3.5. It again becomes negative for $\rho = 0.15$. In the last column, we impose the equal selection on observables and unobservables assumption. In this case, the equal selection is for $\rho = 0.189$. Similarly to the short-run effect, when we impose the equal selection assumption, the estimate of the endogenous variable becomes negative. Again, the zero value lies in between the lower bound of equal selection and the upper bound of no selection. Again, a possible explanation could be that the equal selection assumption is too strong in our case, as we control for many important for the offending behavior variables and we expect a lower selection on unobservables than on observables.

We calculated that a low selection on unobservables relative to the observables can explain the effect of the endogenous variable, in this case it is equal to 0.122. This is, overall, an argument against the long-run effect of victimization on the probability to offend.

Table 3.7 Estimates of victimization before 1995 on offending with different values of ρ

	$\rho = 0$	$\rho = 0.05$	$\rho = 0.1$	$\rho = 0.15$	$\rho = 0.2$	$\rho = \hat{\rho}$ =0.189
Any victimization before 1995	0.231**	0.143	0.054	−0.036	−0.127	−0.103
	[0.092]	[0.092]	[0.092]	[0.092]	[0091]	[0.092]
Marginal effects	**[0.038]**	[0.024]	[0.009]	[-0.006]	[-0.021]	**[-0.018]**

Notes: Robust standard errors are given in brackets[];
** means statistically significant at the 10% level, ** at the 5% level and *** at the 1% level*

To summarize the results, we find no strong evidence of either the short- and long-run effect of victimization on current offending. Though some associations exist for certain samples (especially for the short-run effect of victimization), the control strategy of exploring different samples and Altonji's approach do not support the hypothesis of an effect of past victimization on the offending.

3.6 Conclusion

In this chapter, we analyzed the link between past victimization and the current probability to offend. There are a few theories in social sciences that explain the link between the two but they fail to reach an unanimous conclusion about the magnitude of this relationship, nor do they control for selection bias. We contribute to the literature by extending the correlational findings with the means of econometric techniques. We employ different methods of estimation and use a representative sample from the Netherlands, which contains detailed information about the past offending and victimization history. Using OLS, we find evidence of a short- and long-run effect between victimization and consecutive offending behavior. Namely, past victimization increases the probability to offend with up to 1/3 compared to the baseline average, ceteris paribus. When we focus on groups of the population (those without risky lifestyles, those who do not know the perpetrators of the crime and did not report the crime to the police), we do not find evidence of an association between past victimization and current criminal involvement. After that, we follow the approach by Altonji et al. who use the selection on observables to gauge the selection on unobservables. We find a positive and statistically significant upper bound but the lower bound is not statistically different from zero and is even negative. Finally, the selection on unobservables relative to the selection on observables needs to be only 0.122 (in the case of long-run victimization) to 0.173 (for the short-run victimization) in order for the endogenous variable to explain the whole effect.

Given the high social costs associated with crime, as well as the high victimization rates in the Netherlands (around 25% of the population reports being a victim of a crime), studying in-depth the link between victims and offenders is important. However, in contrast to previous studies, using different methods of estimation, we fail to find evidence that supports an important effect of victimization on the current offending behavior. Therefore, crime prevention policies should not focus on victims of crime; they should rather be given assistance and different resources to overcome the experienced difficulty.

Chapter 4

Do the "young ones" in class need more support? [1]

4.1 Introduction

The age at which students enroll in a given grade might differ due to minimum age school starting laws. These differences can be relatively large, especially at the start of primary education. There is extensive literature on how such relative age assignment affects the in-school performance of students. Many studies find evidence for a short-run effect of relative age. Younger students have, on average, worse test scores than their older peers, and such findings are established for different countries and populations (see, for example, Bickel *et al.* (1991) for the US, Bedard & Dhuey (2006) for 17 OECD countries, Fredricksson & Ockert (2013) for Sweden, Muehlenweg & Puhani (2010) for Germany, Stipek (2002) for a review). However, the effect of relative age on long-run outcomes is less clear. Whereas some studies find a persistent long-term effect of the school starting age on education completion and earnings (Fredricksson & Ockert (2013)), others do not reach such a conclusion (Black *et al.* (2011)). The educational system in place could play a role in whether the relative age effect is increased or attenuated across time. For instance, studies show that in countries where students are streamed into different ability tracks early, the youngest students are disproportionately less likely to be assigned to an academic track (Fredricksson & Ockert (2013), Muehlenweg & Puhani (2010)), likely aggravating the relative age effect. The Netherlands, the country we investigate in our

[1] This chapter is based on Misheva (2015a)

analysis, also has an early tracking system in place. It is therefore important to better understand how the relative age effect works and whether the educational system in place re-enforces the relative age effect or successfully reduces it.

In this paper, we investigate the persistence of the relative age effect across primary education, using longitudinal data from a large sample of students in Dutch primary education. We follow the gaps in test scores between the younger and older students in grades 2 (when they are about 6 years old), 4, 6, and 8 (when pupils are about 12 years old), their results on a high-stakes exam and the advice they receive from the school for the track of secondary education. The main difficulty in investigating the relative age effect is that assignment into grades is not random and the weaker students – who are often relatively younger – are more likely to be retained or sent to special education. As in Bedard and Dhuey (2006), we use the assigned relative age as an instrument for the observed age since the distribution of birth dates is arguably exogenous. In addition, we differentiate between the relative age effect for different samples within our population. We separately look at males versus females, and pupils from different socio-economic backgrounds. This is important because students from higher socio-economic background likely have at their disposal resources which can help them catch up to their better performing peers, such as parental investment, private tutors, more books at home, etc.

We are further contributing to the literature by investigating which policy measures are successful in targeting the relative age effect. Different educational systems have different means to cope with the inequality of their students. In the Netherlands, poor performing students receive remedial teaching, are sometimes retained to repeat a grade, or are sent to special education. We investigate (using the longitudinal nature of our data) whether these factors matter for the relative age effect, and more importantly, which factor is most successful in targeting the performance gap due to the relative age.

We find that in grade 2 the oldest student have an advantage over the younger ones of 9.3 points on the math test and 8.2 points on the language test (or about 0.9 standard deviations). By grade 8, these differences have decreased to 2.6 points for both math and

language (all test scores have been normalized to a mean of 50 and a standard deviation of 10), which translated to about 0.3 standard deviations. There are no substantial differences between boys and girls, but we find differences based on the socio-economic background of students. Namely, for students with a low socio-economic background, the gap by grade 8 is 74% higher than for students with a regular socio-economic background. Younger students perform slightly worse on the high-stakes test at the end of primary education and are less frequently sent to a higher track. Interestingly, teachers have low expectations that the young students will continue in a high-ability track as early as grade 2. We find that the decrease in the performance gap between youngest and oldest pupils across time could be due to receiving remedial teaching, being held back/sent to special education, or maturity, i.e. the younger students catch up to the older ones. The maturity effect is the strongest of the factors, though all of them matter.

The chapter is organized in the following way. In Section 4.2 we discuss previous studies on the topic, in Section 4.3 we explain our estimation strategy. Section 4.4 presents our data and summary statistics, in Section 4.5 we describe our main findings. In Section 4.6 we investigate different hypotheses that could explain our results, in Section 4.7 we decompose the different factors that could explain our findings, and Section 4.8 concludes.

4.2 Previous studies

The relative age effect has been studied for different populations and countries. While we can imagine that in early grades the older students could perform better than the younger ones, it is less clear what will happen in later grades. In principle, if the performance gap were short-lived, the relative age assignment within a grade would not be problematic. Therefore, a key question is how long this difference lasts and whether it depends on the educational system. Given the diversity of educational systems and policies, perhaps some are better at diminishing such inequality among students than others.

A study that addresses both of these issues to some extent is a paper by Bedard and Dhuey (2006), who use data from fourth and eight graders across a number of OECD countries. They find that the youngest students tend to perform worse than the oldest students at both grade levels (a difference that varies from 4-12 percentage points at grade 4 to 2-9 points at grade 8). Not only is the eight-grade difference smaller than the fourth-grade one, but it is not found across all of the countries in their sample. In particular, there is no such gap in Finland and Denmark, which the authors explain with the late tracking and later school starting age, respectively. Therefore, the relative age effect could indeed be affected by the educational system in place.

Muehlenweg and Puhani (2010) focus on Germany, where students are streamed into an academic and a non-academic track as early as the age of 10. They show that relatively younger students are two-thirds as likely to be assigned to the academic track. Deferring tracking to the age of 12 does not mitigate this situation.

Another example is a study by Fredriksson and Ockert (2013), in which the authors use data from the entire Swedish population born between 1935 and 1955. They look at the long-run effects of school starting age on outcomes such as educational attainment and earnings. Higher school starting age implies early advantage, which translates to an increase in the educational attainment by 0.16 years. For the core groups, there is no significant difference of the school starting age on lifetime earnings. Further, the authors find that early disadvantages are exacerbated by early tracking.

Some studies try to separate the age-at-test effect from the relative age effect using information from tests that were taken when students were about the same age. One such example is a paper by Crawford *et al.* (2014), in which they use data from the UK to exploit differences in performance at nationally administered tests and others taken when the pupils were relatively at the same age. They argue that the age at test accounts for the difference in performance between the oldest and youngest students whereas the effect of relative age is close to zero. A second example is a study by Black *et al.* (2011) who use Norwegian data to argue that the long-run effects of school starting age are modest.

There is no long-term effect on educational attainment or earnings (for men). However, the test these authors use was taken just after the participants turned 18, at which point some of them have already left the educational system.

The school starting age could have multifarious effects beyond the scope of test scores. For example, it could affect the probability to receive remedial teaching, be retained, or the teacher's expectations about whether the student would be successful or not. Teachers' expectations are important as many studies demonstrate that lower teacher expectations are associated with lower later achievement of the students (Eder (1981), Downey & Pribesch (2004), Cornwell *et al.* (2013)).

Related to the expectations placed on younger students and their probability to receive remedial teaching, some studies show that the youngest pupils have higher probability to be diagnosed for learning disorders. In one such study, Elder & Lubotsky (2009) argue that the relationship between achievement and kindergarten starting age is primarily driven by the skills older children acquired prior to kindergarten. However, being older decreases the probability to repeat a grade and reduces the probability to be diagnosed for ADHD by around 75 percent of the baseline diagnosis rate. Younger students, though positively affected by having older peers, are also more likely to be diagnosed for a learning disorder. The interpretation the authors give is that school and parental decisions for referral to behavioral professionals are partly based on student's performance or behavior relative to his or her classmates.

Dhuey & Lipscomb (2010) confirm the findings of Elder & Lubotsky (2009). They show that an additional month of relative age decreases the likelihood of receiving special education services by 2 to 5 percent. Such effects are strong for learning disabilities but not for other disabilities. The authors interpret their findings with the fact that disability assessments do not screen for the fact that relatively young students are over-referred for evaluation.

Similar to the latter two papers is a study by Evans *et al.* (2010). They focus on the effect of relative age on the probability to be diagnosed with an ADHD disorder.

Since ADHD is an underlying neurological disorder which incidence should not change dramatically from one birth date to the other, one would not expect to find a difference in the prevalence rate in children born a few days apart. However, the authors find in their study that the relatively younger students (born before the cutoff) have higher probability to be diagnosed for an ADHD. The authors explain the finding with difference in maturity and behavior between the oldest and the youngest. However, they cannot rule out that the problem could be different, namely instead of overdiagnosis of the youngest pupils, there could be underdiagnosis of the oldest ones.

4.3 Empirical strategy

Our main estimation approach is similar to the model used by Bedard and Dhuey (2006). We start with a simple linear model of the following form:

$$T_{igt} = \alpha_1 + \alpha_2 A_{igt} + \alpha_3 X_{igt} + u_{igt} \tag{4.1}$$

Where T_{igt} stands for test score of student i in grade g at time t (or it can capture the probability to be retained or receive remedial teaching), A_{igt} is the observed age of student i in grade g and time t, X_{igt} is a vector of controls, and u_{igt} is an error term. The coefficient of interest α_2 would capture the effect of age on the outcome if there was no selection bias. However, not all students are on track. As we see from our data (descriptive statistics in Table 4.1), relatively younger students tend to be retained more frequently, and older students are sometimes accelerated. Since students who are retained (and have to repeat a grade, for example) have lower ability, OLS estimates would be downward biased.

We continue with an instrumental variables estimation using assigned relative age as an instrument for the observed age. The cut-off date in Dutch primary education is 1^{st} of October. In this way, students born immediately before it (i.e., in September), have

an assigned relative age of 1, those born in August have a relative age of 2, and so forth, until those born in October, who are the oldest, have an assigned relative age of 12. The first-stage equation then becomes:

$$A_{igt} = \beta_1 + \beta_2 R_{igt} + \beta_3 X_{igt} + e_{igt} \tag{4.2}$$

Where all the variables are defined as above, and R_{igt} is the relative age of student i in grade g at wave t. In order to have a valid instrument, first of all it must be correlated with the endogenous variable. In general, students proceed through the education system in line with the assignment rule. Thus, the assigned relative age is clearly an important determinant of the observed age. Second, the exclusion restriction must hold, that is the only effect of the instrument on the outcome is through the endogenous variable. This assumption is violated if ability differs across the months of birth (Bedard & Dhuey (2006)). If the season of birth has no strong effect on the ability, then the instrument is valid. Some studies argue that the in-utero exposure to weather or illness – which are related with the season of birth – could explain later health outcomes, but would not affect the underlying abilities of students (Gortmaker *et al.* (1997), Sham *et al.* (1994), Almond (2006)). They suggest that the birth date is correlated with school entry but not other attributes (Hoogerheide *et al.* (2007)). However, other studies establish a relationship between the season of birth and later outcomes, mostly driven by different fertility patterns across socioeconomic groups, with winter births being disproportionally realized by teenagers and unmarried women (Buckels & Hungerman (2013)). The socioeconomic status of the mother could be related to the parental investment in the skills of the pupil, for example. However, in our case this is not likely to be very problematic because the relatively youngest pupils are summer-born. In addition, in our data we find no evidence that pupils with highly educated parents are more frequently born in the winter in comparison to pupils from disadvantaged background, i.e. there is no systematic targeting of winter or summer months based on the education of the parents.

The second-stage equation becomes:

$$T_{igt} = \gamma_1 + \gamma_2 \hat{A}_{igt} + \gamma_3 X_{igt} + \epsilon_{igt} \tag{4.3}$$

Finally, the reduced-form equation is:

$$T_{igt} = \delta_1 + \delta_2 R_{igt} + \delta_3 X_{igt} + \vartheta_{igt} \tag{4.4}$$

The reduced-form captures the net effect of assigned age on the outcome of interest. The reduced form estimate is expected to be lower than the IV if there is no perfect compliance with the assignment of the treatment. The IV coefficient captures the effect of relative age for the sample of students that comply with the assignment rule. Hence, students for which the observed age would increase when their assigned age would increase. The reduced form then captures the within-grade difference, but including the effect of pupils who have been retained (who have received more schooling and are performing better than if they would be on track).

Finally, we look at the change in the relative age effect across grades. To this aim, we pool all grades and estimate a model similar to the above, but including an interaction term between the age and the grade. The first and second-stage equations are:

$$A_{igt} = \eta_1 + \eta_2 R_{igt} + \eta_3 X_{igt} + \eta_4 R_{igt} * Grade_{it} + \varepsilon_{igt} \tag{4.5}$$

$$T_{igt} = \theta_1 + \theta_2 \hat{A}_{igt} + \theta_3 X_{igt} + \theta_4 \hat{A}_{igt} * Grade_{it} + \nu_{igt} \tag{4.6}$$

Again, we instrument the observed age with the assigned age. Since we observe younger students being retained and this retention tends to increase across the grades, we expect the size of the interaction effect between age and grade to increase with grade.

Though some studies try to distinguish between the age at test, the school starting age and the time spent at school, our focus is on the combined effect of the relative age effect and the factors that matter for it. For instance, if the observed test score gap between the oldest and the youngest students diminishes over time, this could be due

to the younger students catching up (maturity effect), or to a selection bias (the worse performing students are retained or send to special education). In Section 4.6 of this chapter we analyze which factors matter for the change of the relative age effect across grades. The Dutch educational system has a number of measures to target poor performing students, such as remedial classes, retention, and sending students to special education. We investigate which measures are most important for the change in the relative age effect. Are there differences in the effectiveness of the different measures when we focus on pupils from different socio-economic background? In Section 4.7 we investigate these questions.

4.4 Data and the Dutch educational system

In this section we describe our data, provide some descriptions of the Dutch educational system and present summary statistics.

Our data come from the so-called Prima study. The data samples are representative of students in elementary school of the Dutch educational system from grades 2, 4, 6, and 8. The first wave was collected in 1994/1995 and following that, the data have been consecutively collected every 2 years. The last wave of the Prima (Prima 6) was collected in 2004. Each wave has, on average, between 50,000 and 60,000 observations. Overall, we have more than 330,000 observations when we pool all the waves together. Not all the outcomes are observed across all the waves (such as whether the student was receiving remedial teaching or the teacher's expectation of what track the student will be assigned to). We have several test score outcomes. First, all students took tests in math and language in each of the grades given by the Prima administration. The language test score ranges between 800 and 1261 for all the waves and grades, and the math score ranges between 600 and 1361 in Prima 1 and 2 and between 0 and 160 henceforth. Thus, we always use as an outcome a standardized – by year and grade – test score (with a mean 50 and a standard deviation of 10). Students from grade 8 took another test as well, the so called Cito test (short from Cito Eindtoets Basisonderwijs), which is an exam taken at

the end of primary education. It is a comprehensive test, examining students on a number of skills and the score ranges from 500 to 550 points. Again, for ease of interpretation and comparability, we standardize it to a mean of 50 and a standard deviation of 10. The score from the Cito test is important for the track of secondary education in which the student can enroll. The higher the score on the Cito test, the higher the track in which the student can enroll.

In the Dutch educational system, students are tracked after grade 8, which is at the age of 12. There are 3 basic different tracks: VMBO (vocational), HAVO (general secondary) and VWO (pre-university) secondary education. Within the vocational track, there are four different tracks but in our data we can distinguish between three.

As explanatory variables, we include the gender of the student and the weighting factor. The weighting factor is a factor used for the financing schools receive from the government for each student. A pupil with a weighting factor of 1, for example, is a regular (not disadvantaged) native Dutch pupil. The highest weighting factor is 1.90. Students with such a weighting factor have ethnic minority background and low educated parents. Hence, the weighting factor is a proxy for the socio-economic background.

We start by presenting summary statistics in Table 4.1. The first two panels of Table 4.1 show the average standardized test scores by assigned age for each grade. We see that the youngest students have on average lower scores than the older students for all grades and tests. In grade 2, the difference between the oldest and the youngest students is about 5 points on the math and 4 points on the language test. By grade 8, these differences have reduced to about 1 point on the math and 0.8 points on the language test. All the test scores have been standardized to a mean of 50 and a standard deviation of 10.

The third panel of Table 4.1 shows the difference between the observed age of a student born in a given month who is enrolled in a particular grade minus the age of an on-track student (i.e., one complying with the enrollment rule) born in the respective month and from that particular grade. We see that in grade 2, the youngest students are about 3.5 months older than they would be if they were all on track. This gap increases to about

6 months by grade 8, showing that across grades, more of the youngest students are held back. At the same time, the oldest students are close to being on track. The oldest ones (with relative age of 12), are in fact on average younger than what they would be if they were on track, which implies that some have been accelerated. Overall, students with a relative age of 10 and 11 are closest to being on track (we observe smallest difference for them).

Table 4.1 Summary statistics by relative age, PRIMA 1-6

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Month of birth	Sept	Aug	July	June	May	April	Mar	Febr	Jan	Dec	Nov	Oct
Assig-ned age	1	2	3	4	5	6	7	8	9	10	11	12
Average standardized test score in math (Prima 1-6)												
Grade 2	47.80	47.72	48.20	48.57	49.12	49.73	50.37	50.96	51.70	52.45	52.88	52.44
	[9.15]	[9.08]	[9.62]	[9.30]	[9.59]	[9.78]	[9.78]	[10.01]	[10.50]	[10.50]	[10.66]	[10.30]
Grade 4	49.31	49.98	49.33	49.26	49.32	49.68	50.10	50.39	50.70	50.85	51.33	51.05
	[9.60]	[9.59]	[9.71]	[9.82]	[9.92]	[9.91]	[10.22]	[10.12]	[10.19]	[10.20]	[10.16]	[10.21]
Grade 6	49.46	49.45	49.56	49.47	49.43	49.65	50.12	50.31	50.46	50.72	51.15	50.66
	[9.68]	[9.69]	[9.86]	[10.02]	[9.76]	[9.89]	[9.96]	[10.11]	[10.14]	[10.15]	[10.26]	[9.93]
Grade 8	49.71	49.68	49.44	49.48	49.90	49.93	50.09	50.08	50.30	50.41	50.70	50.68
	[9.65]	[9.85]	[9.83]	[9.91]	[9.82]	[9.91]	[10.02]	[9.99]	[10.05]	[10.25]	[10.25]	[9.98]
Average standardized language test scores (Prima 1-6)												
Grade 2	48.13	47.90	48.45	48.59	49.29	49.90	50.32	50.99	51.38	52.02	52.48	52.25
	[9.39]	[9.27]	[9.57]	[9.59]	[9.82]	[9.92]	[10.00]	[10.07]	[10.22]	[10.24]	[10.32]	[10.30]
Grade 4	49.72	49.51	49.47	49.38	49.42	49.79	49.95	50.03	50.49	50.77	50.86	50.86
	[9.86]	[9.64]	[9.83]	[10.09]	[9.99]	[10.08]	[9.96]	[10.05]	[10.07]	[10.06]	[10.14]	[10.02]
Grade 6	49.68	49.56	49.62	49.29	49.42	49.81	50.14	50.27	50.30	50.66	50.82	50.66
	[9.71]	[9.63]	[9.92]	[9.90]	[9.81]	[9.92]	[10.04]	[10.18]	[10.23]	[10.10]	[10.27]	[10.13]
Grade 8	49.72	49.63	49.40	49.52	49.64	49.85	50.04	50.28	50.27	50.60	50.80	50.50
	[9.76]	[9.68]	[9.91]	[9.96]	[9.80]	[10.05]	[10.04]	[9.97]	[10.15]	[10.23]	[10.27]	[10.08]
Difference between observed age and age of on-track student (in months)												
All grades	5.11	4.12	3.70	3.18	2.82	2.32	1.94	1.76	1.51	0.79	0.25	−1.22
	[6.80]	[6.54]	[6.34]	[5.83]	[5.65]	[5.37]	[5.12]	[4.85]	[4.91]	[4.55]	[4.72]	[5.93]
Grade 2	3.48	2.74	2.36	1.99	1.57	1.12	0.96	0.70	0.50	−0.09	−0.58	−2.07
	[5.73]	[5.18]	[5.20]	[4.73]	[4.39]	[4.12]	[4.06]	[3.38]	[3.61]	[3.54]	[3.54]	[5.23]
Grade 4	5.34	4.33	3.81	3.33	2.99	2.47	2.01	1.82	1.57	0.88	0.30	−1.23
	[6.75]	[6.60]	[6.41]	[5.79]	[5.60]	[5.53]	[4.98]	[4.79]	[4.86]	[4.42]	[4.76]	[5.85]
Grade 6	5.98	4.78	4.35	3.72	3.48	2.97	2.40	2.23	2.04	1.25	0.63	−0.77
	[7.19]	[7.01]	[6.80]	[6.27]	[6.14]	[5.84]	[5.52]	[5.43]	[5.24]	[5.07]	[5.18]	[6.08]
Grade 8	6.05	4.90	4.51	3.80	3.37	2.81	2.46	2.37	1.97	1.16	0.68	−0.78
	[7.32]	[7.14]	[6.72]	[6.34]	[6.20]	[5.68]	[5.67]	[5.45]	[5.44]	[4.94]	[5.16]	[6.43]

Note: Observed age is calculated in months, standard deviations are given in brackets; Test scores are standardized to a mean of 50 and standard deviation of 10

Figures 4.1 and 4.2 shows graphically the difference in math and language score across

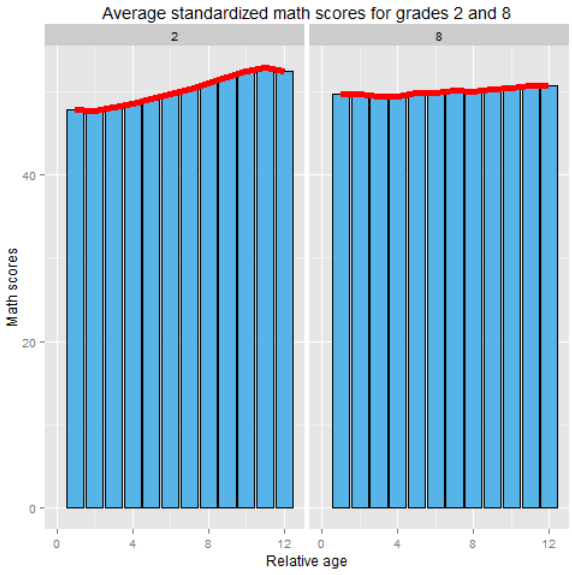


Figure 4.1: Math scores for grade 2 (left) and grade 8 (right)

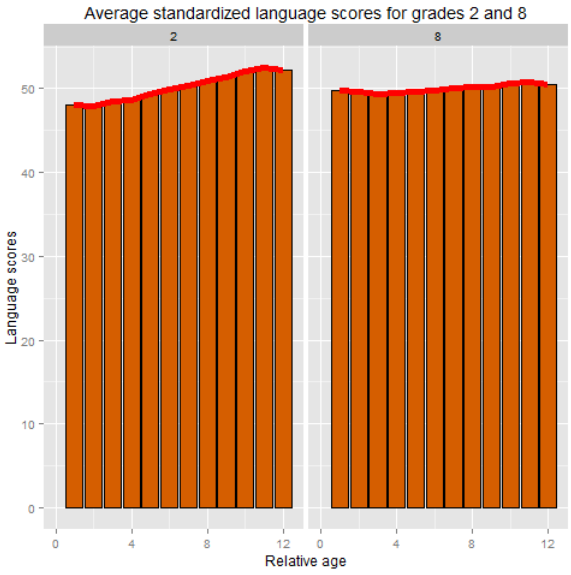


Figure 4.2: Language scores for grade2 (left) and grade 8(right)

assigned relative age for students in grade 2 (left panels) and grade 8 (right panels). The figures demonstrate the pattern we already have seen in Table 4.1. We observe a larger difference in performance between the oldest and the youngest in grade 2 than in grade 8. There is a gap of about 5 points between the oldest and the youngest in grade 2 and of about 1 point in grade 8.

4.5 Effect of relative age on test scores in primary school

We start by exploring the effect of relative age on test scores. In Table 4.2 we present our main estimation results, where we pool all the waves and show the estimates of the effect of age on math and language test scores using OLS, IV and a reduced-form model. In the first column for both of the tests, we see that the coefficient of observed age is positive in grade 2 and is negative henceforth. As argued in the previous section, many students have been retained and they are older, leading to a downward bias in the OLS estimates. As we already have seen in Table 4.1, the proportion of retained students increases with the grade.

Columns (2) and (3) of both panels show the first-stage and the IV estimate. The observed age has been instrumented with the assigned relative age. The first stage is highly significant across all grades and it reduces over grades, in accord with higher failure to comply with the assignment due to the higher retention across grades (see Table 4.1). The IV estimate is statistically significant and is highest in grade 2. To put the coefficients into perspective, the math estimate of 0.85 for grade 2 translates into a 9.3 points difference for the oldest students in comparison with the youngest (on a test with a mean of 50 and a standard deviation of 10), hence, about 0.9 standard deviations. For the language test, this gap in grade 2 is about 8.2 points higher for the oldest students in comparison to the youngest. By grade 8, these gaps have reduced to 2.6 points difference on both tests, or about 0.3 standard deviations.

The reduced form models show the net effect of relative age on the test scores. The IV captures the age effect for the compliers, i.e. for those whose observed age would increase with 1 month with 1 month increase in the relative age. The reduced form includes the effect of students who do not comply and have been retained, for instance, resulting in smaller coefficient than the IV one. Similar to the IV estimates, the reduced-form effect decreases from grades 2 to grades 8, but remains statistically significant at the 1% level. Thus, though it decreases, the effect of relative age has not disappeared by grade 8.

Table 4.2 The effect of relative age on test scores by grade (Prima 1-6)

	Math					**Language**				
	(1) OLS	(2) First stage	(3) IV	(4) Reduced form	N	(1) OLS	(2) First stage	(3) IV	(4) Reduced form	N
Grade 2	0.310^{***} [0.007]	0.605^{***} [0.004]	0.851^{***} [0.016]	0.515^{***} [0.006]	81463	0.277^{***} [0.063]	0.603^{***} [0.004]	0.743^{***} [0.016]	0.448^{***} [0.009]	80875
Grade 4	-0.074^{***} [0.006]	0.505^{***} [0.006]	0.440^{***} [0.020]	0.222^{***} [0.010]	79480	-0.094^{***} [0.006]	0.505^{***} [0.006]	0.302^{***} [0.019]	0.152^{***} [0.009]	79638
Grade 6	-0.250^{***} [0.006]	0.494^{***} [0.006]	0.324^{***} [0.022]	0.160^{***} [0.010]	72334	-0.254^{***} [0.005]	0.491^{***} [0.006]	0.273^{***} [0.021]	0.134^{***} [0.010]	74176
Grade 8	-0.317^{***} [0.006]	0.488^{***} [0.006]	0.240^{***} [0.022]	0.117^{***} [0.010]	68767	-0.285^{***} [0.006]	0.484^{***} [0.006]	0.237^{***} [0.021]	0.115^{***} [0.010]	71236

Note: All models control for gender, the weighting factor, and the wave of PRIMA. Language and math test scores are standardized to have a mean of 50 and a sd of 10 (standardized for each wave and grade). Age is calculated in months.

We continue with the model with interactions. Since the grade of the student could be affected by the observed (or assigned) age, we interacted the grade with the age, pooling all the grades together. Given that we see progressively more retention across grades (and lower compliance), we expect an interaction that is negative and increasing (in magnitude) across grades. This is exactly what we see in Table 4.3. The IV coefficient of a difference between the interaction of age and grade 4 and the baseline interaction of age and grade 2 is -0.415 for math scores and -0.441 for language. These differences increase to -0.619 for math in grade 8 and -0.508 for language. Hence, we observe a clear decrease in the effect of relative age during primary education.

Table 4.3 Effect of age on test scores, interacting age and grade Prima (1-6)

	Math				**Language**			
	OLS	First stage	IV	Reduced form	OLS	First stage	IV	Reduced form
Age	0.277^{***}	0.606^{***}	0.856^{***}	0.518^{***}	0.264^{***}	0.700^{***}	0.743^{***}	0.449^{***}
	[0.007]	[0.005]	[0.016]	[0.005]	[0.007]	[0.019]	[0.016]	[0.009]
Age*Grade 4	-0.349^{***}	0.506^{***}	-0.415^{***}	-0.295^{***}	-0.363^{***}	0.506^{***}	-0.441^{***}	-0.295^{***}
	[0.009]	[0.004]	[0.026]	[0.014]	[0.009]	[0.004]	[0.025]	[0.013]
Age*Grade 6	-0.507^{***}	0.495^{***}	-0.531^{***}	-0.358^{***}	-0.507^{***}	0.491^{***}	-0.472^{***}	-0.315^{***}
	[0.009]	[0.004]	[0.027]	[0.014]	[0.008]	[0.004]	[0.026]	[0.013]
Age*Grade 8	-0.566^{***}	0.487^{***}	-0.619^{***}	-0.402^{***}	-0.535^{***}	0.484^{***}	-0.508^{***}	-0.335^{***}
	[0.008]	[0.004]	[0.028]	[0.014]	[0.009]	[0.004]	[0.026]	[0.013]

Note: See Table 4.2

Before we proceed with other outcomes, we explore the difference in performance for different groups of the population. Concretely, we estimate the effect of age on test scores separately for males and females and for students from different socio-economic background determined by the weighting factor. Table 4.4 shows the effect of age on test scores separately by gender. Both panels are very similar to our estimations in Table 4.2. Overall, there are no substantial differences in the effect of age on the test scores for males versus females. Thus, we do not find evidence of the school starting age affecting males and females differently.

Table 4.4. Estimates of the effect of age on test scores, by gender (Prima 1-6)

Panel A. Estimates for males

	Math					Language				
	OLS	First stage	IV	Reduced form	N	OLS	First stage	IV	Reduced form	N
Grade 2	0.320*** [0.009]	0.578*** [0.006]	0.842*** [0.024]	0.487*** [0.013]	41858	0.291*** [0.009]	0.577*** [0.006]	0.718*** [0.013]	0.414*** [0.013]	41594
Grade 4	−0.079*** [0.008]	0.469*** [0.008]	0.424*** [0.030]	0.199*** [0.014]	39854	−0.093*** [0.008]	0.473*** [0.008]	0.263*** [0.028]	0.124*** [0.013]	39901
Grade 6	−0.249*** [0.008]	0.471*** [0.009]	0.280*** [0.031]	0.132*** [0.014]	36216	−0.252*** [0.008]	0.468*** [0.009]	0.233*** [0.030]	0.109*** [0.014]	36964
Grade 8	−0.314*** [0.008]	0.449*** [0.009]	0.245*** [0.035]	0.110*** [0.015]	34472	−0.282*** [0.008]	0.449*** [0.009]	0.244*** [0.019]	0.110*** [0.014]	35399

Panel B. Estimates for females

	Math					Language				
	OLS	First-stage	IV	Reduced form	N	OLS	First-stage	IV	Reduced form	N
Grade 2	0.299*** [0.010]	0.634*** [0.006]	0.860*** [0.022]	0.545*** [0.007]	39605	0.262*** [0.006]	0.631*** [0.006]	0.767*** [0.022]	0.484*** [0.013]	39356
Grade 4	−0.070*** [0.008]	0.540*** [0.008]	0.455*** [0.026]	0.246*** [0.014]	39627	−0.096*** [0.008]	0.537*** [0.008]	0.336*** [0.025]	0.180*** [0.013]	39694
Grade 6	−0.251*** [0.009]	0.517*** [0.009]	0.364*** [0.030]	0.188*** [0.015]	36117	−0.256*** [0.008]	0.513*** [0.008]	0.308*** [0.028]	0.158*** [0.014]	37214
Grade 8	−0.320*** [0.008]	0.527*** [0.009]	0.236*** [0.029]	0.124*** [0.015]	34298	−0.288*** [0.008]	0.519*** [0.009]	0.231*** [0.028]	0.120*** [0.014]	35673

Note: See Table 4.2

Table 4.5 distinguishes between students with a low and regular socio-economic background, determined by the weighting factor. The weighting factor is a proxy for the socio-economic background of the student (see Section 4.4). The results in Panel A are for students who are not disadvantaged (weighting factor of 1.00) and in Panel B – those with a low socio-economic background (weighting factor of 1.90). The results in Panel A are quite similar to these so far. In grade 2, 1 month increase in the relative age is associated with a 0.9 points increase in the math score, according to the IV estimations. This translates to 9.7 points advantage for 11 months difference in the relative age (0.9 standard deviations). By grade 8, the math score advantage decreases to 2.3 points, again following our IV estimations so far. In Panel B, the coefficients for grade 2 are similar to those from Panel A but over time, they are larger. For instance, the IV coefficient for the math score in grade 8 for this group is 0.363, which implies that the pupils with relative

age of 12 score about 4 points higher than pupils with a relative age of 1. In Panel A, that difference was only 2.3 points, or a 74% increase from Panel A to Panel B. Therefore, for students with a low socio-economic background the gap between the oldest and the youngest students reduces less over time.

Table 4.5. Estimates for effect of age on the weighting factor

Panel A. Weighting factor=1										
	Math					**Language**				
	OLS	First stage	IV	Reduced form	N	OLS	First stage	IV	Reduced form	N
Grade 2	0.386*** [0.011]	0.676*** [0.005]	0.882*** [0.021]	0.596*** [0.013]	43038	0.334*** [0.011]	0.674*** [0.005]	0.739*** [0.020]	0.498*** [0.013]	42698
Grade 4	−0.100*** [0.009]	0.573*** [0.007]	0.432*** [0.025]	0.248*** [0.014]	41000	−0.086*** [0.0095]	0.571*** [0.007]	0.290*** [0.023]	0.166*** [0.013]	41055
Grade 6	−0.289*** [0.009]	0.574*** [0.007]	0.308*** [0.026]	0.177*** [0.014]	36802	−0.290*** [0.009]	0.573*** [0.007]	0.246*** [0.026]	0.141*** [0.014]	37556
Grade 8	−0.387*** [0.004]	0.572*** [0.007]	0.207*** [0.010]	0.118*** [0.015]	34390	−0.331*** [0.010]	0.568*** [0.007]	0.204*** [0.027]	0.116*** [0.015]	35330
Panel B. Weighting factor=1.90										
	Math					**Language**				
	OLS	First stage	IV	Reduced form	N	OLS	First stage	IV	Reduced form	N
Grade 2	0.246*** [0.011]	0.486*** [0.010]	0.780*** [0.037]	0.379*** [0.009]	21562	0.227*** [0.006]	0.479*** [0.010]	0.725*** [0.036]	0.352*** [0.017]	21435
Grade 4	−0.044*** [0.0094]	0.365*** [0.013]	0.467*** [0.054]	0.170*** [0.018]	20500	−0.105*** [0.009]	0.367*** [0.013]	0.350*** [0.051]	0.128*** [0.018]	20510
Grade 6	−0.194*** [0.010]	0.331*** [0.016]	0.356*** [0.068]	0.118*** [0.021]	17465	−0.214*** [0.008]	0.329*** [0.016]	0.304*** [0.058]	0.100*** [0.018]	18188
Grade 8	−0.239*** [0.010]	0.306*** [0.017]	0.363*** [0.078]	0.111*** [0.022]	15909	−0.243*** [0.008]	0.303*** [0.017]	0.375*** [0.069]	0.114*** [0.019]	16597

Note: See Table 4.2

It is important to explore whether the age effect is present also for the Cito test as it is a high-stakes test, which determines the type of secondary education the students are assigned to. The estimation results are given in Table 4.6. Since there are some missing observations for the Cito scores (and also due to the merging of the cito data sets with those with students information), the Cito sample is smaller than the entire sample of eight-graders. For comparison of this sample to the whole population, in Panels B and C of Table 4.6, we present estimates for math and language scores for this smaller sample,

which we label the Cito sample. The estimates from Panels B and C are quite similar to those for the eight-graders from Table 4.2. Panel A shows the estimates for the effect of age on the Cito score. The OLS again has a negative sign, indicating a downward bias. The first-stage is very similar to the first-stage for grade 8 from Table 4.2. Overall, about 48% of pupils are on track (i.e., complying with the treatment). The IV estimate of 0.325 translates into a 3.6 points difference for 11 months of relative age (that is, the difference between the oldest and the youngest students), again for a test with mean of 50 and a standard deviation of 10; or about 0.36 standard deviations. Thus, there is a significant gap in the test scores between the oldest and the youngest students for the high-stakes test.

Table 4.6 Effect of age on CITO scores, math and language for CITO samples (Prima 1-6)

	OLS	First-stage	IV	Reduced form	N
Panel A. Cito scores					
Grade 8	-0.359^{***}	0.487^{***}	0.325^{***}	0.158^{***}	47717
	[0.007]	[0.008]	[0.027]	[0.012]	
Panel B. Math test scores for the CITO sample					
Grade 8	-0.311^{***}	0.491^{***}	0.263^{***}	0.129^{***}	44743
	[0.007]	[0.008]	[0.027]	[0.013]	
Panel C. Language scores for the CITO sample					
Grade 8	-0.271^{***}	0.489^{***}	0.241^{***}	0.118^{***}	46125
	[0.007]	[0.008]	[0.026]	[0.012]	

Note: See Table 4.2

Next, we look at the effect of age on the advice for the secondary education track and the expectations of the teachers on whether the students will continue to a high-ability track. In Panel A of Table 4.7, we show teachers' expectations about whether the student will go to a high-ability track. This information was only collected for Prima waves 4 to 6, which explains our smaller samples. We see again a significant and persistent effect across grades. For students in grade 8, for example, the teacher expects that the oldest students in comparison to the youngest are about half a point more likely to go to high track secondary education on a scale from 1 to 5 (with 1 very small chances and 5 very high chances). In grade 2, that gap is equal to 0.78 points (on the scale of 1 to 5).

The advice the oldest students receive in comparison to the youngest ones has a coefficient of 0.047 in the IV estimate. This is equivalent to about half a point difference between the oldest and youngest students, or about 50% higher chance for the older students to be assigned to vwo (pre-university) instead of havo (general secondary) track. It is also interesting to note that the teachers' expectations come very close to the advice the students receive.

Table 4.7 Effect of age on track of secondary education

Panel A. Teacher's expectation the student will go to high track education

	OLS	First-stage	IV	Reduced form	N
Grade 2	-0.014^{***} [0.001]	0.615^{***} [0.007]	0.071^{***} [0.003]	0.044^{***} [0.002]	25325
Grade 4	-0.040^{***} [0.001]	0.501^{***} [0.010]	0.046^{***} [0.004]	0.023^{***} [0.002]	22877
Grade 6	-0.047^{***} [0.001]	0.471^{***} [0.011]	0.047^{***} [0.005]	0.022^{***} [0.002]	21968
Grade 8	-0.052^{***} [0.001]	0.470^{***} [0.011]	0.042^{***} [0.005]	0.020^{***} [0.002]	21637
Panel B. Advice students received from the school, PRIMA data					
	OLS	First-stage	IV	Reduced form	N
Grade 8	-0.053^{***} [0.001]	0.485^{***} [0.008]	0.047^{***} [0.004]	0.023^{***} [0.002]	46172

Note: In Panel A, teacher expectation of whether the student will continue in a higher track in secondary education ranges from 1 to 5, higher values stand for bigger chance. Models control for gender, dummies for weighting factor, and for wave of the Prima.

In Panel B, advice and track range from 1 to 6 (1 is for vmbo lwoo/bl, 6 is vwo).

4.6 Why does the relative age effect decline?

We have shown so far that the relatively younger students perform on average worse in comparison to the their older peers but by grade 8 the gap is smaller than it is at grade 2 (though still highly significant).

The Dutch educational system has a number of policy measures in place to help poor performing students. Some are given extra help in the form of remedial teaching, others (for whom the remedial teaching has not worked, for instance) are retained and repeat

a grade, or are re-directed to special education. However, the change could also be the result of growing older. In this section, we explore different explanations for this decrease.

4.6.1 Probability of retention

One possible explanation for the reduction in the test score gap over time is a selection effect. The younger students who are performing the worst could be held back and only the more capable are allowed to proceed to the next grade, resulting in a smaller gap than the one which we would observe without retention. In the following table we present the estimates of the effect of age on the probability of being retained. Our outcome – the probability to be retained – is a binary variable coded with 0 if the student is on track and with 1 if he/she is older than an on-track student in the respective grade and wave. Therefore, it will capture not only the probability of being held back at each grade but the accumulated retention rate over grades, including starting school later. About 29% of the youngest students in grade 2 are older than they should be if they were on track, and only 1.3% of the oldest. By grade 8, around 46% of the pupils with a relative age of 1 are older than they would be if they were on track (or coded as 1 by our retention variable), and 9% of the oldest students.

The OLS results in the first column are positive, which could again be due to a bias. The IV and reduced form estimates are negative and significant for all the grades. In grade 2, one month increase in the relative age is associated with a 2.4% decrease in the probability of being held back according to our reduced-form results, which increases to 3% in grade 4. Interestingly, after grade 4 the probability changes very little. As noted above, our retention variable does not perfectly isolate students who repeat a grade as it also captures students who started later. Even so, the magnitude increases over time, demonstrating that the youngest students are held back more frequently than the older ones, which could be another reason for the decrease in the test scores gap.

Table 4.8 Effect of age on probability to be retained

	OLS	First-stage	IV	Reduced form	N
Grade 2	0.044^{***}	0.585^{***}	-0.042^{***}	-0.024^{***}	74258
	[0.0002]	[0.005]	[0.001]	[0.0003]	
Grade 4	0.051^{***}	0.496^{***}	-0.061^{***}	-0.030^{****}	70144
	[0.00001]	[0.006]	[0.002]	[0.0004]	
Grade 6	0.052^{***}	0.490^{***}	-0.063^{***}	-0.031^{***}	77563
	[0.0001]	[0.007]	[0.002]	[0.0004]	
Grade 8	0.053^{***}	0.474^{***}	-0.067^{***}	-0.031^{***}	62517
	[0.0002]	[0.007]	[0.002]	[0.0004]	

Note: See Table 4.2

4.6.2 Probability of being observed in the next wave

Another explanation could be that the youngest students are not only more frequently retained but are sent to special education, for example. Here, we examine the probability to observe a student in the next wave given that he/she was in the previous one. The results are given in Table 4.9.

We see that the effect of relative age on the probability to observe a student conditional on observing her in the previous wave is positive and significant in the first two rows, and is not significant in the third row. The IV estimate in the first row shows that one month increase in the relative age is associated with a 1.1 percentage point increase in the probability to observe the student two years later. In the second row the effect is smaller though still significant. There are a number of explanations for why we do not observe a student in two consecutive waves, such as the student is held back, sent to special education, has moved and changed the school.

Table 4.9 Probability to be observed in consecutive waves

	OLS	First-stage	IV	Reduced form	N
Grade 2-4	0.025^{***}	0.629^{***}	0.011^{***}	0.006^{***}	67072
	[0.0002]	[0.008]	[0.001]	[0.0005]	
Grade 4-6	0.019^{***}	0.510^{***}	0.004^{***}	0.002^{***}	64367
	[0.0002]	[0.009]	[0.001]	[0.0005]	
Grade 6-8	0.016^{***}	0.499^{***}	0.001	0.0002	59945
	[0.0002]	[0.009]	[0.001]	[0.001]	

Note: See Table 4.2

4.6.3 Remedial teaching

Poor performing students are likely to receive extra help in the form of additional classes and assistance, more attention, perhaps specific exercises, etc. Thus, another explanation for a reduction in the gap between the oldest and youngest students over time is remedial teaching, targeted at the poorest performing students.

In Table 4.10 we show the effect of age on the likelihood of receiving remedial teaching. We see that across all grades, the increase in relative age is associated with a decrease in the likelihood to receive remedial teaching; the strongest evidence is found for students in grade 2, which decreases almost threefold by grade 8, though it is still statistically significant.

Table 4.10 Effect of age on receiving remedial teaching, Prima (4-6)

	OLS	First-stage	IV	Reduced form	N
Grade 2	0.012^{***}	0.616^{***}	-0.071^{***}	-0.044^{***}	27166
	[0.001]	[0.007]	[0.003]	[0.002]	
Grade 4	0.033^{***}	0.499^{***}	-0.033^{***}	-0.017^{***}	23233
	[0.001]	[0.010]	[0.004]	[0.002]	
Grade 6	0.038^{***}	0.473^{***}	-0.039^{***}	-0.018^{***}	22162
	[0.001]	[0.011]	[0.005]	[0.002]	
Grade 8	0.041^{***}	0.472^{***}	-0.025^{***}	-0.012^{***}	21653
	[0.001]	[0.011]	[0.004]	[0.002]	

Note: See Table 4.2

4.6.4 Maturity effect

A final explanation for our findings could be that the decrease in the performance gap

over time is due to a maturity effect, i.e. the youngest students catching up with their older peers. We investigate whether this is indeed the case.

First, we use the longitudinal nature of the Prima data to estimate the effect of age on changes in the test scores. For example, in the second wave there are students who were also interviewed in the previous wave. If a student is in grade 2 when the first survey was collected, he should be in grade 4 at the time of the second survey. That is, of course, if he does not fall behind (is retained), is not sent to special education, if he is not accelerated and does not change schools, and the school does not leave the sample. We explore how the relative age affects the difference in the test scores. The equation we estimate is of the following form:

$$T_{i,\,g+2,\,t+2} - T_{igt} = \delta_1 + \delta_2 X_{igt} + \delta_3 A_{igt} + \rho_{igt} \tag{4.7}$$

Where T_{igt} is the test score of a student i in grade$g$ at time t and $T_{i,\,g+2,\,t+2}$ is his/her test score in grade $g+2$ at time $t+2$. The other variables have already been defined. Again, we instrument the observed age A with the assigned age R.

The results are given in Table 4.11. Each row shows the test score difference between two consecutive grades, i.e. grade 4 and grade 2, 6 and 4 and 8 and 6. We also control for remedial teaching, thus the outcome will be net of retention and remedial teaching and due to the maturity effect. Because we have information for remedial teaching only for waves 4 to 6, the samples become smaller. The IV estimate in the first row for math scores is -0.303. This means that the increase in relative age is associated with a decrease in the test score difference between grades 4 and 2. In other words, the oldest students compared to the youngest ones have about 3.3 points lower grade 4-grade 2 math score difference (tests standardized to a mean of 50 and a standard deviation of 10), or about 0.33 standard deviations. The oldest students score on average higher than the youngest ones across all grades, and a negative difference means that the youngest ones are increasing their average test scores over time more than the oldest ones are. The result does not mean that the youngest students are outperforming the oldest ones.

Table 4.11 Effect of age on across-grade difference in test scores, PRIMA (1-6)

	Math					Language				
	OLS	First stage	IV	Reduced form	N	OLS	First stage	IV	Reduced form	N
Grade 4-2	-0.323^{***} [0.022]	0.599^{***} [0.038]	-0.303^{***} [0.046]	-0.181^{***} [0.017]	12248	-0.304^{***} [0.023]	0.606^{***} [0.010]	-0.381^{***} [0.029]	-0.352^{***} [0.0489]	11969
Grade 6-4	-0.177^{***} [0.015]	0.516^{***} [0.047]	-0.120^{***} [0.0457]	-0.062^{***} [0.023]	11567	-0.115^{***} [0.018]	0.520^{***} [0.013]	-0.107^{**} [0.053]	-0.056^{**} [0.028]	11898
Grade 8-6	-0.075^{***} [0.0148]	0.523^{***} [0.014]	-0.119^{***} [0.026]	-0.062^{***} [0.022]	11242	0.014 [0.016]	0.524^{***} [0.014]	-0.084^{*} [0.048]	-0.044^{*} [0.025]	12002

Note: See Table 4.2

Second, we look at the test score progress across grades for students who are more frequently on track (the two relatively oldest quarters of pupils). Since these students are less often held back, there is a lower selection bias for this group; thus if there is a decrease in the test score gap over time, it will be mostly due to maturity. Table 4.12 shows the results from such an exercise. There is a positive and significant increase in test scores associated with the increase in the relative age in grade 2. The coefficients decrease henceforth but remain statistically significant. The coefficients in Table 4.12 are overall higher than those in Table 4.2, confirming once more that retenion and being sent to special education contribute to the decrease in the test scores gap across grades.

Table 4.12 Effect of age on test scores for the 2 oldest quarter of pupils, PRIMA (1-6)

	Math					Language				
	OLS	First stage	IV	Reduced form	N	OLS	First stage	IV	Reduced form	N
Grade 2	0.180^{***} [0.012]	0.549^{***} [0.009]	0.921^{***} [0.043]	0.506^{***} [0.023]	44463	0.163^{***} [0.018]	0.546^{***} [0.009]	1.05^{***} [0.273]	0.761^{***} [0.041]	44231
Grade 4	-0.127^{***} [0.009]	0.461^{***} [0.011]	0.545^{***} [0.052]	0.251^{**} [0.022]	45379	-0.135^{***} [0.085]	0.455^{***} [0.011]	0.410^{***} [0.049]	0.415^{***} [0.022]	45450
Grade 6	-0.313^{***} [0.009]	0.444^{***} [0.013]	0.431^{***} [0.056]	0.192^{***} [0.023]	41592	-0.306^{***} [0.008]	0.444^{***} [0.013]	0.327^{***} [0.053]	0.145^{***} [0.022]	42584
Grade 8	-0.358^{***} [0.009]	0.464^{***} [0.013]	0.329^{***} [0.054]	0.153^{**} [0.024]	39390	-0.327^{***} [0.008]	0.461^{***} [0.013]	0.306^{***} [0.052]	0.141^{***} [0.023]	40714

Note: See Table 4.2

To sum up our findings in this section, the decreasing gap in the test scores over time

between the relatively older and relatively younger students could be due to a number of factors: maturity (the youngest students catch up with the older ones), retaining younger students or sending them to special education, and receiving remedial teaching. We showed that, in our case, all these factors seem to matter. In the next section, we decompose the decrease of the test-scores gap to obtain an approximation of the importance of the different factors.

4.7 Decomposition of the decreasing test score gap

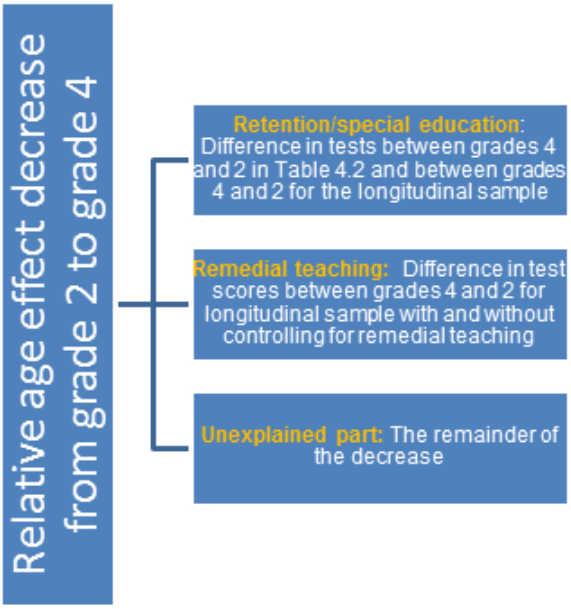
In the previous section we saw that different factors contribute to the decrease in the test score between the oldest and the youngest students across grades. In this section, we try to decompose the different effects. The decomposition we perform aims to approximate the relevant importance of each of the factors.

For the decomposition, we exploit the longitudinal nature of the data. First, for finding out the imporantance of retention or sending students to special education, we compare the decrease in the relative age effect over grades in our main estimates in Table 4.2 with a similar decrease while using a longitudinal sample, i.e. pupils we observe in consecutive waves. For them, a potential decrease in the relative age effect cannot be attributed to retention or sending students to special education. Hence, for them, the decrease of the relative age effect can only be attributed to the maturity effect or to receiving remedial teaching. Therefore, by comparing the across-grade relative age decrease in Table 4.2 with the decrease in the panel sample, we can gauge the importance of retention/sending pupils to special education.

Second, for finding out the importance of remedial teaching, we compare the across-grade difference for students we observe in the sample for two consecutive waves with and without controlling for remedial teaching. Finally, the remainder in the decrease of the relative age effect across grades is an unexplained part, which could be due to maturity, but could also include other factors. Figure 2 summarizes this explanation using the decomposition in the decrease of the relative age effect between grades 2 and 4 as an

example.

Figure 4.2. Chart demonstrating the decomposition



4.7.1 Decrease from grade 2 to grade 4

In Table 4.2 (our main estimates), the estimated effect of the relative age on the math score decreases from 0.851 (the IV coefficient) in grade 2 to 0.440 in grade 4 (a decrease in the coefficient of 0.41), and the language test score decreases from 0.740 to 0.300, or a reduction of about 0.440. Below, we assess which part of this 0.41 or 0.440 decrease is due to each of the factors.

Table 4.13 shows the effect of the age on test scores for pupils in grades 2 and 4 only for the panel sample, i.e. students observed in grade 2 who were also present in grade 4. Thus, the decreasing effect over time for this group then cannot be the result of retention or sending students to special education, but can only be the result of maturity and/or remedial teaching. In Table 4.13 the decrease in math is from 0.750 to 0.400, and for language – from 0.640 to 0.290; hence a decrease of 0.350 for each test. Therefore, this decrease is smaller than the decrease for the total sample, which also includes retaining students/sending them to special education. Taking the difference in the decrease between the coefficients in Table 4.2 and in Table 4.13 will give us an approximation of the part

of the effect which is **due to retention/special education**. For the math test, this is equal to the difference between 0.41 and 0.35, or 0.06; for the language, the difference between 0.44 and 0.350, or 0.09.

Table 4.13 Effect of age on test scores for grade 2 and 4, PRIMA (1-6)

	Math					Language				
	OLS	First stage	IV	Reduced form	N	OLS	First stage	IV	Reduced form	N
Grade 2	0.243^{***} [0.011]	0.586^{***} [0.006]	0.750^{***} [0.024]	0.439^{***} [0.014]	35511	0.205^{***} [0.010]	0.586^{***} [0.006]	0.638^{***} [0.024]	0.373^{***} [0.093]	35340
Grade 4	-0.083^{***} [0.009]	0.514^{***} [0.008]	0.399^{***} [0.028]	0.205^{***} [0.014]	36842	-0.105^{***} [0.008]	0.516^{***} [0.008]	0.288^{***} [0.026]	0.149^{***} [0.013]	36999

In Table 4.14, we take the difference in test scores between grades 4 and 2 for students we observe in consecutive waves. We control for remedial teaching. In Panel B we estimate the same sample but do not control for remedial teaching. Overall, the estimates when including remedial teaching are higher than when we do not account for it, indicating that the remedial teaching is effective in reducing the gap between the oldest and the youngest students. The difference in the coefficients for math is about 0.001, and for language – about 0.04. Therefore, **the contribution of remedial teaching** in the decrease in the gap from grade 2 to grade 4 is 0.001 for math and 0.04 for language.

Table 4.14 Panel A:Effect of age on test scores for grades 2 and 4, controlling for remedial teaching PRIMA (4-6)

	Math					Language				
	OLS	First stage	IV	Reduced form	N	OLS	First stage	IV	Reduced form	N
Grade 4- 2	-0.323^{***} [0.013]	0.605^{***} [0.010]	-0.303^{***} [0.046]	-0.181^{***} [0.028]	12240	$-0.304 * **$ [0.0213]	0.606^{***} [0.010]	-0.352^{***} [0.048]	-0.212^{***} [0.029]	11969

Panel B: Effect of age on test scores for grades 2 and 4, PRIMA (4-6), no control for remedial teaching

	OLS	First stage	IV	Reduced form	N	OLS	First stage	IV	Reduced form	N
Grade 4-2	-0.327^{***} [0.013]	0.586^{***} [0.010]	-0.304^{***} [0.048]	-0.176^{***} [0.028]	12240	-0.304^{***} [0.023]	0.588^{***} [0.010]	-0.398^{***} [0.049]	-0.212^{***} [0.029]	11969

For clarity, we summarize the decomposition results in Table 4.15. The decrease in the gap between the oldest and youngest students could be decomposed in the following way. Retention/sending students to special education contributes with 0.06 to 0.09 to the decrease, remedial teaching with up to 0.04 for the language score and 0.001 for the math, and the rest of the decrease in the coefficient is unexplained, or what we attribute to a time effect (the younger students catching up with their older peers); in this case, it is about 85.4% for the math test decrease and about 70% for the language one.

Table 4.15 Decomposition in the test score decrease between grades 4 and 2

	Math	Language
Total $\triangle$ (grade 4-2)	0.41	0.44
Retention/special education	0.06	0.09
Remedial teaching	0.001	0.04
Unexplained	0.349	0.31

4.7.2 Decomposition of the test score decrease from grade 6 to 4, and from 8 to 6

In this section, we perform the same analysis as from the decrease in test scores between grades 2 and 4. In Table 4.16 we present the decomposition for the change from grades 4 to 6 (Panel A), and from 6 to 8 (Panel B). The complete set of tables, following those in Section 4.7.1 are given in Appendix 1. The unexplained part is most often the largest. Retention/sending pupils to special education is the second most important factor when we look at the decrease between grades 6 and 4. It is less important in the decrease of the relative age effect btween grades 8 and 6.

Table 4.16 Decomposition for the decrease in test scores between grades 6 and 4, and 8 and 6

Panel A. Decomposition for the decrease between grades 6 and 4		
	Math	Language
Total $\triangle$ (grade 6-4)	0.116	0.029
Retention/special education	0.037	0.012
Remedial teaching	0.012	0.012
Unexplained	0.067	0.005
Panel B. Decomposition for the decrease between grades 8 and 6		
	Math	Language
Total $\triangle$ (grade 8-6)	0.084	0.036
Retention/special education	0.016	0.005
Remedial teaching	0.027	0.006
Unexplained	0.041	0.025

4.7.3 Decomposition for students from different socio-economic backgrounds

We showed in Section 4.5 that for students with a low socio-economic background, the gap between oldest and youngest pupils in grade 8 is 74% larger than for regular students. In this part, we repeat the decompositions while distinguishing between students from different socio-economic background, as determined by the weighting factor. The results are given in Table 4.17, Panles A to C. For students who do not come from a disadvantaged background, the unexplained part – which we attribute to time effect – is the largest in all of the tables. Interestingly, for students from a low socio-economic background, retention/sending students to special education and receiving remedial teaching are also

important factors (their importance varies based on the subject and grade).

Table 4.17. Decomposition for the decrease in test scores across grades by weighting factor

Panel A. Decomposition in the test score decrease between grades 4 and 2				
	Weighting factor=1.00		**Weighting factor=1.90**	
	Math	Language	Math	Language
Total △ (grade 4 -grade 2)	0.45	0.45	0.313	0.425
Retention/special education	0.098	0.116	-0.023	0.121
Remedial teaching	0.003	0.044	0.001	0.148
Unexplained part	0.349	0.29	0.273	0.256
Panel B. Decomposition in the test score decrease between grades 6 and 4				
	Weighting factor=1.00		**Weighting factor=1.90**	
	Math	Language	Math	Language
Total △ (grade 4 -grade 2)	0.124	0.048	0.111	0.046
Retention/special education	0.038	0.018	0.067	0.016
Remedial teaching	0.011	0.024	0.005	0.009
Unexplained part	0.075	0.006	0.039	0.021
Panel C. Decomposition in the test score decrease between grades 8 and 6				
	Weighting factor=1.00		**Weighting factor=1.90**	
	Math	Language	Math	Language
Total △ (grade 4 -grade 2)	0.101	0.042	0.007	0.071
Retention/special education	0.028	0.013	0.021	0
Remedial teaching	0.034	0.025	0.032	0.017
Unexplained part	0.039	0.004	*	0.054

4.8 Conclusion

School starting rules can place students born a few days apart in different grades. Thus, some students are 11 months older than their peers in the same grade. This difference is substantial in the early grades. It is less clear how such a difference progresses in later classes. In this chapter, we contribute to the literature on the relative age effect by investigating the relative age effect for primary school students in the Netherlands and decomposing the significance of different factors that affect the change of this effect across grades.

We find that in grade 2 the oldest students score about 9.3 points higher than their younger classmates on a math test and 8.2 higher on a language test (on a test with a mean of 50 and a standard deviation of 10). By grade 8, these differences decrease to 2.6 points difference in favor of the older students for both math and language tests. These

differences are also observed in the high-stakes test pupils take at the end of primary education, where older pupils score about 3.6 points higher than their younger classmates. Younger students are less likely to go to a higher track of education, and their teachers less frequently expect them to continue in a high-track of secondary education. There is not a substantial difference in the effect of age on the test score gap over time between boys and girls. Interestingly, for students with a low socio-economic background, the gap has not reduced as much by grade 8, and for them we find a difference of about 4 points in favor of the older students.

An important question given our findings is what drives the decrease in test scores between the oldest and the youngest students. We investigate a number of hypotheses. First of all, the poor performing students, who are more frequently relatively younger, receive remedial teaching to help them with their difficulties. They often could be retained to repeat a grade or sent to special education. Alternatively, it could be due to a maturity effect, i.e. the youngest students catching up with their older peers. We find that all these factors matter. We decompose the different channels and find that the maturity effect is the largest factor across all grades.

Given that the relative age still has an effect on a high-stakes test in grade 8 and on the assignment into a secondary education track, education policies need to address this problem more adequately – such as additional assistance and remedial teaching targeting younger students (especially those from a low socio-economic background) or reforming the tests (so that students take them when they are at a specific age). Another option is to take the age of a student into account when pupils take the Cito test, or keep the age in mind when giving advice for the track of secondary education. Alternatively, classes could be split into into a few groups based on the half year (or quarter) in which pupils are born. Of course, this will only be applicable when schools have enough students. The early tracking, present in the Dutch educational system, is likely to enhance the gap between older and younger students as by the beginning of secondary education it has clearly not disappeared. This provides an argument for additional policies to address the performance

gap between the older and younger students.

Appendix

Table A1. Effect of age on test scores for grade 4 and 6, panel sample PRIMA (1-6)

	Math					Language				
	OLS	First stage	IV	Reduced form	N	OLS	First stage	IV	Reduced form	N
Grade 4	-0.083^{***}	0.514^{***}	0.399^{***}	0.205^{***}	36842	-0.105^{***}	0.516^{***}	0.288^{***}	0.149^{***}	36999
	[0.009]	[0.008]	[0.028]	[0.014]		[0.008]	[0.008]	[0.026]	[0.013]	
Grade 6	-0.250^{***}	0.500^{***}	0.321^{***}	0.161^{***}	36811	-0.195^{***}	0.501^{***}	0.276^{***}	0.173^{***}	37682
	[0.008]	[0.0114]	[0.030]	[0.014]		[0.009]	[0.009]	[0.031]	[0.0154]	

Table A2. A. Effect of age on test scores for grades 4 and 6, controlling for remedial teaching PRIMA (4-6)

	Math					Language				
	OLS	First stage	IV	Reduced form	N	OLS	First stage	IV	Reduced form	N
Grade 6- 4	-0.177^{***}	0.516^{***}	-0.120^{**}	-0.062^{**}	11564	-0.115^{***}	0.520^{***}	-0.107^{***}	-0.056^{**}	11903
	[0.015]	[0.013]	[0.045]	[0.023]		[0.018]	[0.013]	[0.053]	[0.028]	

Panel B. Effect of age on test scores for grades 2 and 4, PRIMA (4-6), no control for remedial teaching

	Math					Language				
	OLS	First stage	IV	Reduced form	N	OLS	First stage	IV	Reduced form	N
Grade 6-4	-0.165^{****}	0.501^{***}	-0.132^{***}	-0.066^{***}	11564	-0.126^{***}	0.504^{***}	-0.095^{*}	-0.048^{*}	11903
	[0.015]	[0.013]	[0.046]	[0.017]		[0.018]	[0.013]	[0.055]	[0.028]	

Table A3. Effect of age on test scores for grade 6 and 8, quasi-panel sample PRIMA (1-6)

	Math					Language				
	OLS	First stage	IV	Reduced form	N	OLS	First stage	IV	Reduced form	N
Grade 6	-0.250^{***}	0.500^{***}	0.321^{***}	0.161^{***}	36811	-0.195^{***}	0.501^{***}	0.276^{***}	0.137^{***}	37682
	[0.008]	[0.0114]	[0.030]	[0.014]		[0.009]	[0.009]	[0.031]	[0.015]	
Grade 8	-0.239^{***}	0.497^{***}	0.294^{***}	0.146^{***}	22417	-0.245^{***}	0.501^{***}	0.271^{***}	0.135^{***}	22996
	[0.011]	[0.011]	[0.038]	[0.0184]		[0.0108]	[0.011]	[0.028]	[0.0174]	

Table A4. A. Effect of age on test scores for grades 6 and 8, controlling for remedial teaching PRIMA (4-6)

	Math					Language				
	OLS	First stage	IV	Reduced form	N	OLS	First stage	IV	Reduced form	N
Grade 8- 6	-0.075^{***}	0.523^{***}	-0.119^{**}	-0.062^{**}	11243	0.014	0.524^{***}	-0.084^{*}	-0.044^{*}	11903
	[0.014]	[0.051]	[0.0425]	[0.0223]		[0.015]	[0.014]	[0.053]	[0.028]	

Panel B. Effect of age on test scores for grades 2 and 4, PRIMA (4-6), no control for remedial teaching

	OLS	First stage	IV	Reduced form	N	OLS	First stage	IV	Reduced form	N
Grade 8-6	-0.041^{**}	0.502^{***}	-0.155^{***}	-0.078^{***}	11243	0.037^{**}	0.505^{***}	-0.106^{**}	-0.054^{**}	22366
	[0.014]	[0.014]	[0.044]	[0.017]		[0.015]	[0.014]	[0.050]	[0.025]	

Chapter 5

What determines emotional well-being? The role of traumatic events: Evidence using twin data[1]

5.1 Introduction

Though happiness has interested humans for centuries, only in recent decades have economists abandoned their firm belief that economic agents reveal their preferences solely through their choices. Before the reign of the decision utility theories, economists believed that utility incorporated one's evaluation of their experiences of pleasures and pains. Edgeworth postulated that we could grasp someone's evaluation of their experiences, and do so on a cardinal scale. Vilfred Pareto, however, disagreed that utility can be measured cardinally and for the next few decades economic theory largely abandoned the idea of measuring someone's well-being by directly asking them about their experiences. Things changed when Richard Easterlin conducted a seminal study in 1974 (Easterlin (1974)), which demonstrated that growth in US income was not supplemented by growth in happiness. This revived economists' interest in well-being.

Happiness, satisfaction with life and subjective well-being are typically used interchangeably in economic studies mainly because the concepts are often confounded (Kahneman & Deaton (2010)). This paper will focus on emotional well-being. Emotional well-being is usually defined as the emotional quality of everyday experiences, the positive

[1]This chapter is based on V. Misheva (2015b), Journal of Happiness Studies, online since 6th October. The paper has greatly benefited from suggestions by Dinand Webbink, the participants at the ESE brownbag seminars, the EHERO seminars, and at the Social Sciences Promotion Centre conference in Porto (2014).

and negative affect that makes one's life pleasant or unpleasant (Kahneman & Deaton (2010)). In contrast to Kahneman and Deaton, we can only measure emotional well-being with a single self-reported question. It falls under the affective component of life evaluation (Veenhoven (2009)). Veenhoven (2009) argues that the hedonic level of affect is a less problematic measure because it does not require a subjective evaluation of how well one feels. On the other hand, contentment with one's life is a deliberate cognitive process. As such, it requires assessment of one's quality of life according to his chosen criteria (i.e., how life is compared to how life should be). Whether the different concepts defining quality of life are interchangeable is established most clearly by comparing their determinants. Does income increase life satisfaction but not happiness and emotional well-being? Do other factors similarly affect different measures of subjective well-being? This paper will conduct a validity check of the stylized facts in the literature concerning the determinants of subjective well-being.

Studying emotional well-being and its determinants is important not only because of the intrinsic interest of happiness as an ultimate goal, but also because of the relevance of well-being for economic policy. In order to make a Pareto-improving change, policymakers should be aware of its potential effects in terms of individual well-being, or utility. Further, well-being could provide information for the overall condition of institutions (Frey & Stutzer (2002b)), such as rule of law, state of the government, etc. Lastly, it allows us to capture human well-being directly and permits us to directly test the validity of economic and psychological theories.

We employ data on Australian twins to perform our analysis. A number of twin studies in the field of happiness literature use their genetic similarities to evaluate the heritability of well-being. Tellegen *et al.* (1988) and Lykken & Tellegen (1996) maintain that common family environment does not significantly impact personality traits and subjective well-being but that genes have a large effect. Interestingly, the authors find that monozygotic twins rared together and monozygotic twins rared apart display heritability of their well-being of around 0.8 and unshared environment must account for the remaining 20% of

the variance in the well-being. However, these authors employed rather small samples, so their estimates should be viewed with caution. Using a nationally representative sample of twins from the US, Neve *et al.* (2012) revealed that genetic variation explains around 33% of the variation in life satisfaction. Relying on a sample of Dutch adolescent twins and four different measures of well-being, Bartels & Boomsma (2009) found that there are underlying additive genetic and non-genetic factors that cause clustering in the measures of well-being. They found that the heritability of subjective well-being (SWB) ranges from 40 to 50%.

These studies used twin data to test the hypothesis that happiness is a genetically determined trait. We also tested to what extent monozygotic versus dizygotic twins provide similar responses to the well-being question. Most importantly, we employ a twin fixed effects strategy, which to the best of our knowledge has not been used in previous studies of well-being. With such a strategy, all the unobserved common for the twins characteristics are removed even if they cannot be measured. This is likely to reduce the selectivity bias. Therefore, twin-fixed effect models potentially surpass correlational analysis.

Furthermore, our study represents the first attempt in economics (to the best of our knowledge) to test the validity of the adaptation hypothesis using a large number of traumatic events and applying econometric techniques. To test the validity of this theory, we will analyze the effect of a number of traumatic events that occurred in adulthood and more recently.

This chapter is organized as follows. Section 5.2 reviews the stylized facts of the well-being literature, providing a motivation for the factors we include in our analysis. Section 5.3 describes the empirical strategy, its benefits and potential limitations. In Section 5.4, we describe the data, provide some descriptive statistics and evidence for within-twin variation. In Section 5.5 we present our estimation results, testing of the robustness of previous literature and provide evidence for the adaptation hypothesis. In Section 5.6 we test the robustness of our results, and Section 5.7 concludes.

5.2 Determinants of the emotional well-being

In this section we review the stylized facts from the literature and explain what controls we include in our analysis. We explain the self-reported emotional well-being by three separate groups of covariates: life situation, life history and life abilities.

5.2.1 Life situation

Almost all studies of subjective well-being account for the age of the respondent. However, the relationship between age and well-being is not straightforward. Some studies show that the well-being decreases with age – such as Shields *et al.* (2009) who use Australian data, and Peiro (2006) who finds a negative relationship for the majority of the 16 nations he investigates, whereas others find more of U-shaped relationship with lowest point around middle-age (Ferreri-Carbonell & Frijters (2002)). A possible explanation is that age could be a proxy for unobservables and the relationship is highly dependent on the regressors included.

Having children is either not statistically significant, or it has a negative effect on one's well-being. Shields *et al.* (2009) find the number of children decreases the life satisfaction for both men and women. Bjornskov *et al.* (2008) use data from 80 nations to find a negative relationship between having one, two, and three or more children and the life satisfaction. This negative relationship is especially well-established among single or divorced parents (Becchetti & Pelloni (2013)).

Many studies do not find a significant difference in the well-being between the two genders, and some find that males report, on average, higher life satisfaction whereas women report higher job satisfaction (Clark (1997)) .

The majority of studies finds a positive effect of the years of schooling on the self-reported happiness (Kozcan (2013)with data from Germany, Dittmann & Goebel (2010) also using data from West and East Germany, and Graham (2005) with data from the US). However, Krause (2013) in her sample of unemployed respondents finds that the higher

level of education is negatively correlated with the life satisfaction. Given that highly educated people have higher opportunity cost of unemployment, and probably higher aspirations, this finding is not surprising.

A very important factor for the subjective well-being is one's marital status. Studies show that married people (or those cohabiting with a partner) are, on average, happier than single, divorced, or widowed individuals. One might argue that there are selection effects and happy people are more attractive and more likely to be married (Stutzer & Frey (2006)), but longitudinal research (that compares the life satisfaction before and after getting married) establishes that the direction is mainly from marriage to life satisfaction (Coombs (1991), Gove *et al.* (1990), Kessler & Essex (1982),Headey *et al.* (2008)). Also studies that compare married to widowed and/or separated people find that those who are married report much higher life satsfaction to those who no longer are (Dittmann & Goebel (2010), Peiro (2006), Hagerty & Veenhoven (2003)).

Another very good predictor of well-being is the health status (Okun & Stock (1987), Myers & Diener (1995), Helliwell & Putnam (2004), Becchetti & Pelloni (2013), Bohnke & Kohler (2007)). Good health is a strong and positive correlate of self-reported happiness (Dittmann & Goebel (2010), Headey *et al.* (1993), Michalos & Kahlke (2010)). For example, Bruni & Stanca (2006) using waves 2, 3 and 4 of the World Values Survey (and data from 80 nations), establish that a 1 point increase in the self-reported health increases the life satisfaction with 0.5 points (life satisfaction measured on the scale from 1 to 10 and the health status from 1 to 5). Those who are in poor health, have fallen serious ill, or have become handicapped report overall a lower life satisfaction than those who report to be in good health. Nettle (2005) finds that among the population in Great Britain born in one week in March in 1958, those who report physical illness rate their life, on average, with 1 score lower than those in good health (on the 1 to 10 scale). Brooks & McKinlay (1983) look at a sample of severely head injured patients, and according to reports of their close relatives, the average decrease in their happiness after the injury amounts to about 35%. All of these, and numerous other studies, find a strong relationship between the health

status and the subjective well-being. However, just as with the marital status, potential reverse causality could be present. Veenhoven (2008) reviews 30 studies about the effect of happiness on health, and using longetivity as a measure of health, discovers that although happiness does not predict longevity in sick populations, it predicts longevity in healthy populations – and finds a large effect, comparable to whether one is smoking or not.

Religiousness is another, often controlled for variable. Studies find different magnitude and sign of the religion depending on how religousness is measured – simply by defining oneself as being "religious", to belonging to religious organization, attending prayers, etc. Okulicz-Kozaryn (2010), for example, uses the 4th World Values Survey (and data from 70 nations) to find a negative association between belief in God or belonging to a religious denomination and life satisfaction, but a positive one between participation in church activities or attending religious services and the life satisfaction. Halman (1987), using data from 10 EU countries, finds only a small positive relationship between defining oneself as religious and the self-reported happiness. Shields *et al.* (2009) in their study on Australia find a positive but small association between the importance of religion and life satisfaction.

Among the economic factors, the income and being unemployed are among the best predictors of subjective well-being. Studies find that being unemployed decreases well-being considerably (Clark & Oswald (1994), Frey & Stutzer (2002b)). One could again imagine that if a worker is unhappy, he could perform worse at work, is less motivated and therefore, puts less effort, which could lead to a lay-off (explored, for instance, in Krause (2013)). However, the general consensus is that the direction is from unemployment to lower life satisfaction and not the other way around (Clark & Oswald (1994), Frey & Stutzer (2002a)).

Income is among the most interesting factors for economists but the effect of income on well-being is not as straightforward as one might assume. Usually, basic economic utility theory posits that agents are better-off when they have more (in this case more income). However, research shows that high income is not necessarily associated with higher well-being. This is originally known as the "Easterlin paradox" (as income increased

substantially over time in countries like USA and Japan, life satisfaction did not). However, most studies agree that (very) low levels of income are significantly associated with lower happiness and life satisfaction (Easterlin (2003), Diener *et al.* (1999), Frey & Stutzer (2002b)). At high levels of income, the effect is lower. Hagerty & Veenhoven (2003) find that national level of happiness increases with the increase of the GDP but these effects are short-lived for rich countries and most prominent for poor nations (again confirming that once the basic needs and wants of humans are satisfied, life satisfaction does not linearly increase with the increase in income). There is some evidence that higher income could be associated with higher level of distress (Thoits & Hannan (1979)), or it could be due to adaptation and aspiration theories. The theory of adaptation postulates that humans adapt quickly to new circumstances and if there is an increase in their income, they update and increase their aspirations, therefore the gap between aspiration and achievement is not reduced (Michalos (1991)). In short, there are diminishing marginal returns from the increase in income. Lastly, Kahneman and Deaton (2010) distinguish between the effect of income on emotional well-being and on the life satisfaction. They establish that higher income raises the life satisfaction but not the emotional well-being.

5.2.2 Life abilities

Psychological research stresses the importance of personality in evaluating one's life satisfaction and well-being (Diener *et al.* (1999)). One of the theories postulates that individuals have inborn dispositions to be happy or unhappy. This evidence comes from studies such as Lykken and Tellegen (1996) and Tellegen et al. (1988), who find that monozygotic twins raised together or raised apart are much more similar to each other than fraternal twins, who were raised in the same family. Lykken and Tellegen (1996) find that from 40 to 55% of the variation in current subjective well-being can be explained by genes, and around 80% of the long-term well-being is heritable.

Among the personal traits that have received the most empirical and theoretical attention are extroversion and neuroticism (Diener *et al.* (1999)). Extroversion is typically

associated with positive affect (Headey *et al.* (1993)), whereas neuroticism influences negative affect (Costa & McCrae (1980)). Muffels & Kemperman (2011) followed a sample of German women over time and found a large negative coefficient of neuroticism and a smaller and positive coefficient of the extroversion.

5.2.3 Life history

Social science is not unanimous on whether hardships decrease or increase one's happiness. Ventegodt (1999) looks at the effect of early life conditions and traumas on the quality of life 30 years later in Denmark. He finds only relatively weak connections between early life experiences (such as the child was placed in a children's home, mother using drugs for mental illness, child adopted in first year of life) and the quality of life later, concluding that children that survive to adulthood are resilient to many adverse events in early life. Elder (1974) in his study "Children of the Great Depression", establishes that respondents who remember most hardships from the Depression times demonstrate higher increase in happiness in later years, which he explains as having a very low standard for evaluating subsequent life-events.

A number of studies come to an opposite conclusion. One such example is a study by Barschak (1951). In his study, students from countries directly involved in the Second World War reported being significantly less happy than students from countries that were not directly involved in the war. Kainulainen (1998) studies the effect of ever experiencing violent behavior on the satisfaction of life among people in a province in Finland and finds strong and negative correlation not only for victimization that happened the year before the interview, but in any point in life. Brorsson *et al.* (1993) use a sample of the Swedish population, particularly distinguishing people that were involved in traffic accidents that happened two years before the interviews, and find that accident victims score lower not only on positive affect, but also on negative one, indicating reduced emotionality. Angeles (2010) uses a British panel and finds adaptation to separation, health deterioration or

improvement over time. So, overall, the magnitude of the effect of traumatic experiences on the happiness is not clear, nor is it clear how long the effect of a certain traumatic experience would last. We need to point out that the evidence reviewed above stems from correlational studies, which does not allow us to claim that a certain traumatic experience causes a permanent change in the well-being with a respective magnitude.

5.2.4 Covariates that we include

Based on the reviewed literature, we include three categories of factors in our analysis. In the first one we account for the life situation and include **demographic, personal and familial characteristics**. Such variables account for the gender, the age, the marital status, having children, education level of the respondent, religiousness, and the self-reported health. We can explicitly control for some personality traits (life abilities) – whether someone is extroverted, or neurotic – following the short-form revised Eysenck's personality traits questionnaire. In a second category of characteristics we separate the **economic factors**. In that group we have an indicator variable for being unemployed, and 2 variables for reporting income in the highest and lowest quartiles of the income distribution.

The last category of factors includes **traumatic experiences** throughout one's life (life history). To account for traumatic childhood experiences, we include variables for physical abuse (that occurred when the respondent was between 6 and 13), for sexual abuse (either by a family member or an outsider), and for neglect. We also have information about whether the respondent has been arrested, has spent time in jail, or has ever done something he/she could have been arrested for, though they were not. Although criminal behavior is highly endogenous, an experience like spending time in jail could potentially have overlasting effects on one's emotional well-being. Among traumatic experiences in adulthood, we know whether the interviewee has been in an accident, has been through a natural disaster, has been assaulted (which lead to physical injuries), has been raped, has been held captive, has witnessed a serious injury done to someone else (or a murder), or

has a close person who has been through something traumatic.

5.3 Empirical strategy

5.3.1 Estimation framework

First, we start with a simple model where we pool the sample and treat the twin pairs as individual observations in order to compare our results with findings from other studies. We first explain the emotional well-being with different personal characteristics and economic factors, and later add the reports about traumatic events in childhood and in recent years. We focus on a cardinal linear relationship.[2]We estimated an equation of the following form:

$$Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i \tag{5.1}$$

Where Y_i is our measure of well-being and it is equal to 1 if the respondent rates his emotional well-being as "poor", 2, as "fair", 3 as "good", and 4 as "excellent". In our vector of personal characteristics X_i we include our three groups of factors – personal and familial characteristics, economic factors and traumatic experiences; ε_i is an error term.

Since it is quite difficult to imagine an experimental setting where some participants are assigned to experience certain events (especially traumatic ones), and others were not, a simple OLS would fail to allow causal interpretation of our estimates. OLS results would be inflicted by an omitted variables bias stemming from unobserved heterogeneity. To reduce this bias, we proceed with exploiting the twin nature of the data by estimating a twin-fixed effects model. With twin-fixed effects, common family and genetic factors that we do not account for will not be a problem because they are removed. The equation we will estimate would be of the form:

[2] We estimated an ordinal logit model as well but obtained very similar results. Ferrer-i-Carbonell and Frijters (2002) show that whether one assumes cardinality or ordinality would not lead to different results. For ease of interpretation of the coefficients (especially so with the fixed effects) we stick to a cardinal relationship.

$$Y_{ij} = \gamma_1 + \gamma_2 X_{ij} + \mu_j + u_{ij} \tag{5.2}$$

Where Y_{ij} is the self-reported emotional well-being of twin i in family j. We again focus on a cardinal relationship.[3] X_{ij} is a vector of characteristics that vary within the twin pair. μ_j captures the common familial and genetic background and u_{ij} is an error term. Note that dizygotic (fraternal) twins are on average as much as alike as any other siblings and only monozygotic (identical) twins are genetically absolutely the same. Therefore, we will mostly rely on identical twin estimates as they remove the highest part of unobserved heterogeneity.

A central idea of measuring subjective well-being, though not explicitly tested in economic studies, is adaptation. Humans tend to adjust to both good and bad circumstances (Diener *et al.* (1999)). People have different coping mechanisms and the point of reference they use at any point in time could change. However, the evidence for this hedonic treadmill is not unequivocal (see Section 5.2 for more details). Unfortunately, we cannot single out this effect, and by focusing on the the long-run effects of traumatic events, we will inevitably also be capturing the ability of a person to adapt and recover after an adverse experience. Therefore, we differentiate between events in the past 3 years and the past 1 year to check whether there are differences in the way more recent and not so recent events affect the emotional well-being on average.

5.3.2 Advantages and potential concerns

Equation (5.2) will help us alleviate a few existing problems in the well-being literature. First of all, estimating a relationship like (5.2) will reduce a potential endogeneity bias by

[3] No simple transformation is available that will purge the ordered response from the within-pair fixed effects. There are some attempts to consistently estimate an ordered logit (there are no such attempts in the ordered probit), where researchers attempt to collapse J different categories into two classes (see Winkelmann & Winkelmann (1998); Baetschmann *et al.* (2011)), which are later estimated with a conditional maximum likelihood. However, the coefficients derived with such approaches are very difficult to interpret.

controlling for common family background and genetic similarities. With twin-fixed effects the common background and genetic factors are dropped, even if they are correlated with the other covariates. Time-invariant personal covariates would be eliminated with panel data because many important observable characteristics (such as education level, marital status, number of children, gender, etc.) do not vary, or vary very little over time. With our approach such time-invariant characteristics will not be eliminated. However, twin-fixed effects has its own shortcomings, which we will highlight in the following paragraphs.

To be able to claim causality, we need to have random assignment and an experimental design, which is not the case in our study. Clearly, performing the analysis using identical twins alleviates the endogeneity, but does not remove it completely. What is more, one might be concerned that differences in important observed characteristics within twin pairs are not random, which would lead to a bias in our estimation results. For instance, in the case of schooling, if a family is likely to send further the twin who shows promise in his/her education, the estimates of the effect of schooling on the EWB will be biased upwards. But if the family is trying to reduce the inequality and invests more in the worse performing twin, our estimates will be downward biased (Ashenfelter & Krueger (1994)). This could hold for the differences in many of the other variables. Unfortunately, we cannot know in which direction omitted variables may affect our results.

Another potential concern is measurement error. Attenuation caused by a measurement error is increased with twin-fixed effects due to the correlation within the pair, which leads to a lower effect, and biases our results downwards (Ashenfelter & Krueger (1994), Griliches (1979)). Such measurement error is typically augmented by instrumenting the twin's response with the response of his/her co-twin. Unfortunately, we do not have such information for the majority of the variables in our data, and we are unable to adjust our estimates for the size of the measurement error. The threat of a measurement error is perhaps highest in our traumatic experiences measures. The data are retrospective and people might unintentionally repress the memory for some traumatic event, or they might intentionally provide a different answer when inquired about certain experiences. If this

is indeed the case, then our estimates will underestimate the true effect of adverse events on one's emotional well-being, and essentially, we would obtain a lower bound.

Second, a measurement error could stem from our emotional well-being measure. Though a measurement error in the outcome variable will not lead to biased estimates, it will still reduce the precision of out estimates since the measured variance would be higher than the true one. However, there is no universal and unique scale in measuring the (emotional) well-being. One might argue that measuring the well-being is questionable since the way one rates his/her well-being is influenced by their mood during the interview, the experiences they have been through, and by their coping mechanisms. Even the framing of the questions, the order in which they appear, and some events (as external as the weather during the interview day) could potentially have an effect on the provided answers. We will try to address all these potential problems one by one.

First of all, the well-being questions were asked at the beginning of the interview and before all the questions about traumatic events during one's childhood and adulthood. Experimental studies show that participants who were asked to describe a recent sad event and were afterwards asked to value their life satisfaction, gave an overall lower life-satisfaction score than subjects who were urged to remember recent happy events (Schwarz & Clore (1983)). Since our well-being questions precede the traumatic experiences sections, this is not likely to affect the mood of the respondents.

The next potential threat to the precision of our estimates would arise if the current mood affects the provided well-being ratings. This problem was alleviated in several ways. First of all, good or bad mood is in many cases random (perpeniducular to one's characteristics). What is more, studies using repeated observations over time establish that there are no consistent associations between mood states and subjective well-being ratings on average. Test-retest correlations for mood are relatively small, whereas for subjective well-being are substantially larger. Pavot & Diener (1993) found a two-months test-retest correlation of 0.82 for rates of subjective well-being, and Williams *et al.* (1991) reported a correlation of 0.86 on views of life scale.

Furthermore, it is very likely that personality affects the way people rate their well-being. Whether one is a pessimist or optimist, extroverted or introverted, are shown to be correlated with the life satisfaction reports, and such personality traits might predispose individuals to experience positive and negative life events (Wilson (1967), Costa & McCrae (1980), Lucas (2008)). Some psychological studies show that even social desirability is a personality trait that enhances well-being rather than a source of error variance (Diener *et al.* (1991)). We can explicitly control for extroversion and neuroticism using the revised small-scale Eysenck personality trait test, and implicitly for all other genetically determined personality traits. To address the issue of the current mood affecting one's responses during the interview and for the respondent's general pessimism/optimism, we can compare the within-twin differences to questions that should have been answered in the same way and regress these differences on the difference in their emotional well-being rating.

Moreover, we can test for a comparison effect to one's co-twin. Numerous studies stress the importance of comparison to others in evaluating one's well-being. For instance, if the overall unemployment is high, then losing a job would not have such a negative impact on one's happiness (Clark & Oswald (1994)). We cannot account for the effect stemming from all the peers of the respondent but we can use the information provided by the co-twin. For example, if one's twin is employed in a high-paying job, is happily married and enjoys great health, this could act as a negative externality on the co-twin who is not married, has health problems, or is in a low-paying job. In our robustness checks section we include some characteristics of the co-twin as right hand-side variables in the EWB equation.

A usual concern in twin studies is the presence of contamination (or peer) effects. In our case, if one of the twins goes through a traumatic event, and the other not, the second might still be affected by his co-twin's experience. This will not be very problematic for our estimations because we can explicitly control for a traumatic experience that happened to the respondent's co-twin, or anyone else close to him/her.

Finally, we need to be cautious about reverse causality. The presence of selection

effects is especially troublesome for the marital status and the health variables. As we discussed in Section 5.2, some studies argue that happy people have higher probability to get married and/or enjoy better health. To check whether a selection bias is a big problem for our estimates, we use health and marital status information from an earlier wave as instruments for the current ones, and as predictors of the current EWB.

5.4 Data

We use data from the second wave of the so-called Younger Cohort of the Australian Twin Register gathered between 1996-2000. We also include some information from the first wave, collected 1989-1990. Altogether we have 6265 single observations, of which 5530 form complete twin pairs. From them, 2332 are monozygotic twins (1166 pairs) and 3198 are dizygotic twins (1599 pairs).

Our outcome variable is constructed based on the following question:

"How would you describe your emotional well-being? Would you say it is excellent, good, fair or poor? 1-Poor, 2-Fair, 3-Good, 4-Excellent"[4]

This is a rather similar question to the most popular ones in the literature. The Eurobarometer asks respondents how satisfied they are with their lives, on the whole, with answers ranging from "Not at all satisfied" to "Very satisfied". The question is almost identical to the one asked in the European Social (Values) Survey (ESS) and in the German Socio-Economic Panel Survey (GSOEP). The World Values Survey uses a scale from 0 to 10 with 0 being "Dissatisfied/Not very happy" with one's life and 10 being "Satisfied/Happy".

In our sample, only about 14% rate their emotional well-being as Poor or Fair, all the others said it was either Good, or Excellent (and the modal response being 3, i.e. Good). This is in accordance to findings in other studies, which establish that people usually tend

[4] We need to acknowledge that some respondents could misinterpret this question, and instead of their overall assessment of feeling good or not (the affective component of happiness), they could have interpreted it as an inquiry about their mental health, which in itself is a component of happiness but is not equivalent to happiness. Such errors in responding, as well as errors in estimating one's overall well-being are rather common.

to rate their happiness or life satisfaction rather high, or in other words, there is bunching towards the top of the scale (Diener *et al.* (1999), Clark *et al.* (2008)).

Table 5.1 shows the summary of the outcome variable by gender and zygosity of the twins. The rest of the descriptive statistics are given in Appendix 2, Table A1. We see that most respondents rate their emotional well-being rather high, with an average of 3.16. According to the table in Appendix 2, the average level of education is around 12 years, the average age of the respondents is around 30, and 64% of them report to be married or cohabiting with someone at the time the interview took place, while around 7% are divorced or separated. Around 4% report to be unemployed, and almost 70% report to be religious. The self-reported health is, overall, predominantly high. The scale for rating the health is similar to the one used to assess the EWB with 1 standing for having poor health, and 4- excellent health status. Around 34% report physical abuse, and 12% – sexual abuse. From the traumatic experiences, being involved in an accident and seeing someone else being seriously injured or killed are reported by most of the respondents (19 and 23%, respectively). Around 5% report rape, and around 10% have been assaulted.

It is important to compare the prevalence rates of the traumatic events in our sample to the general incidence rates for Australia, at least for those variables for which this is possible. According to statistics by the Australian government, around 17% of women 18 and older, and 4% of the men report sexual assault, in most cases by a perpetrator they knew; around 18% of women report being sexually abused before the age of 16, and around 4% of the women in the sample reported forced intercourse over their lifetime. Various studies of the prevalence of sexual abuse in Australia indicate that the rates range from around 10% (Mamun *et al.* (2007)) to 16% (Dunne *et al.* (2003)) for males and from 12% (Dunne *et al.* (2003)) to 42% (Mazza & Dennerstein (2001)) for females. Thus, our sexual abuse reported by around 12% of the whole population (15% among women only), and 5% of rape (8% among women) are in the same ballpark as these official prevalence rate statistics.

In the second row of Table 5.1, we display the intra-class correlation in the emotional

well-being report between dizygotic and monozygotic twins, obtained with a oneway analysis of variance (ANOVA) using random effects. Put simply, this allows us to determine what portion of the variance in the EWB is due to between-twin difference compared to within-twin difference (i.e. degree of relationship within the twin pair is estimated by the proportion of the total variability that is accounted for by between class variance). We computed that the within-pair correlation for DZ twins is 0.08, and that for MZ twins is 0.24. This means that among MZ twins, around 24% of the overall variation in EWB comes from between twin variation. In general, the larger the intra-class correlation is, the less variation comes from within the pair relative to the means between the pairs. Therefore, MZ twins are much more similar to each other than are DZ twins in their reports of EWB. Similar to findings of other studies (Neve *et al.* (2012), Kohler *et al.* (2005)), we find higher correlation for MZ twins than for DZ ones (our numbers, in fact, come quite close to those in Kohler *et al.* (2005) who find a intra-class correlation of 0.21 for younger MZ twins and 0.24 for older ones). This indicates the presence of genetic dispositions on the variation of EWB, and smaller relevance of the shared environment. Note also that this higher intra-class correlation would exacerbate the presence of any measurement error in the identical twin-fixed effects estimation, reducing the power of our estimates.

Table 5.1. Summary of emotional well-being

	All	DZ twins	MZ twins	Males	Females
Emotional	3.16	3.15	3.19	3.21	3.12
well-being	[0.7]	[0.71]	[0.68]	[0.69]	[0.71]
[1,2,3,4]					
Intra-class		0.08	0.24		
correlation		[0.03]	[0.03]		
Observations	6265	3198	2332	2803	3462

Note: standard deviations are given in brackets []

Those reporting to be in poor versus those in excellent EWB differ in many characteristics.[5] Those with excellent well-being more frequently reported to be married, have higher education, better health; those with poor EWB are more often unemployed, divorced or

[5]Results available from the author upon request.

separated, or have income in the lowest quartile. Interestingly, those with an excellent EWB did not report more frequently income in the highest quartile of the distribution in comparison to those with a poor one. This already signals the weak association between high income and emotional well-being (similar to Kahneman & Deaton (2010)). Furthermore, respondents with poor emotional well-being reported more frequently physical and sexual abuse and neglect, more often have spent time in jail or done something for which they could be arrested. In addition, this group more frequently indicated involvement in an accident, assault, rape, captivity or having someone close to them who had been through a traumatic experience.

For a sample of twins, within-twin variation is especially important, as it enables us to perform our twin-fixed effects estimations. We have analyzed the proportion of families (among the whole sample of twins and among MZ ones) in which the response of one of the twins differs from that of his/her co-twin. We found a high degree of within-pair variation in the variables of interest, especially so in the traumatic events. [6] [7]

5.5 Empirical results

In this section we present estimates of the effect of the different factors on the emotional well-being. First, we present OLS and twin-fixed effects, starting with the stylized facts that affect the well-being (personal and economic ones, see Table 5.2), and then we add the traumatic experiences (Table 5.3). Finally, we focus on traumatic experiences that occurred in the past 3 and past 1 year (combined in Table 5.4).

[6] Results available from the author upon request.

[7] One could be concerned about where this within variation stems from in some of the cases. Whereas traumatic experiences are quite often a negative shock, out of the control of the individual, it is more difficult to justify the difference in the child abuse reports. The sexual abuse very often stems from an outsider and so is again often a negative exogenous event. The physical abuse is an interesting phenomenon and social science studies propose as an explanation a "single child" targeting where the parents would abuse only one of the children in the family. This could be due to some characteristics of the child (his gender, idiosyncratic behavior, physical and mental problems), or to some parental characteristics (mental disorders, abuse of alcohol, drugs, etc.). Some researchers also argue that abusive parents might targer a single child instead of all children in the family in order not to attract much outside attention, or prevent cooperation between the children and their reporting to the authorities (see Jaffee *et al.* (2004)).

5.5.1 Testing some stylized facts

Table 5.2 shows our first set of regression results. In it, we explain the EWB with the common in the literature factors that influence one's well-being. Here, and in all the tables to follow, all the scale variables (EWB, health status, extroversion and neuroticism) have been normalized to have mean of 0 and variance of 1. The standard errors are clustered on the twin-pair level. In the first column we estimate a linear relationship. The results in column (1) are in accordance with what the majority of other studies in the literature find. The EWB decreases with age. One year increase in age reduces the EWB with 0.013 of its standard deviation. Using the GSOEP data Ferreri-Carbonell & Frijters (2002) (henceforth FCF) find that the coefficient for age is -0.03 (well-being measured on the 0-10 scale). Graham (2005) using data from the Latinobarometer finds a decrease of 0.025 of the happiness with increase of age (with an ordered logit), and identical coefficient using data from the US; and Graham *et al.* (2004) use data from Russia to find a negative association between age and happiness of a magnitude of -0.067.

One extra child is associated with 0.03 standard deviations decrease in the EWB. FCF also find a coefficient of -0.03 when using an ordered probit, and -0.05 when using an OLS. Kohler *et al.* (2005) find a small positive effect of having children equal to 0.028 for females, but it disappears and becomes negative when they account for current partnership.

Our marriage variable is highly significant and positive. Being married/cohabiting with someone is associated with 0.27 standard deviations increase in the EWB. In their OLS with controls, FCF obtain a coefficient of 0.23. Using information from the WVS, Helliwell & Putnam (2004) (henceforth HP) find that being married or living with someone increases the self-reported happiness between 0.31 to 0.48 (using a rescaled measure of happiness from 1-4 to a 1-10 scale). Using twin data from Denmark, Kohler *et al.* (2005) find a positive effect of 0.26 of currently being in a relationship on the well-being of females and 0.32 on the well-being of males (and it does not decrease with age). Stutzer & Frey (2006) use the GSOEP data in a study trying to account for selection effects, and find a coefficient of around 0.3 in an OLS estimation on life satisfaction.

A 1 sd increase in the health status is associated with a 0.35 sd increase in the EWB; for FCF the subjective health variable has a coefficient of 0.39, and HP find an effect of health of 0.54 to 0.65 on the happiness and the life satisfaction, respectively. Graham (2004) finds an effect of health equal to almost 0.5 using the Latinobarometer, of 0.46 for Russia, and 0.62 for the US. In our case, being religious does not increase one's happiness. As we saw in Section 2, there is no strong predictive power of religiousness on life satisfaction in Western countries, and the way religiousness is measured also matters. For example, we cannot account for the engagement of the respondent in religious activities, which some studies show is the main aspect associated with increased life satisfaction (Dolan *et al.* (2008)).

Being unemployed and having income in the lowest quartile have a negative effect on the EWB. Our unemployment variable of -0.24 is similar to the 0.33 decrease in the life satisfaction that Tella *et al.* (2001) find, using a scale from 1 to 4. Graham (2004) establishes a negative coefficient of 0.49 for the effect of unemployment on happiness with the Latinobarometer data, one of -0.66 for Russia, of -0.68 for the US. HP find a negative effect of 0.36 of unemployment on happiness, and one equal to -0.65 for life satisfaction. Reporting high income, similarly to what Kahneman & Deaton (2010) find, does not increase the emotional well-being. The EWB also increases with the extroversion score, and decreases with the neuroticism score.

Overall, our OLS results are very similar to findings in the literature, despite the different scales, countries of the studies, years of interviews, and regressors included. However, we want to test the robustness of these findings using twin-fixed effects and purging of the common familial background and genetic similarities. These estimates are shown in the rest of Table 5.2. In the second column, we include the whole sample of twins. We include as explanatory variables only those which differ within the pair. In general, the significance and magnitude of the majority of the variables is preserved as we move from OLS to the twin-fixed effects. This implies that there is little bias stemming from shared environment.

In the last column we present the estimates for monozygotic twins only. The coefficient of being married is almost identical to the one from the OLS estimation in column (1). Reporting divorce/separation is no longer statistically significant, though the coefficient is comparable in magnitude to the first column. The years of education variable increases in magnitude compared to columns (1) and (2), and now one year increase in the years of schooling is associated with a 0.06 standard deviations increase of the EWB. The self-reported health continues to be statistically significant at the 1% level. Now, neither of the variables capturing one's extroversion or neuroticism are statistically significant, which could indicate that they are to a large extent genetically determined. Being unemployed is no longer statistically significant but reporting income in the lowest quartile increases in absolute value, though only significant at the 10% level. The low income variable could be a proxy for the unemployment and this could explain why we no longer see any separate effect of the unemployment status.

Testing the stylized facts from the literature in Table 5.2, we can already draw some conclusions. The marital status, years of education, health status and having low income continue to be significant determinants of the EWB when we apply twin-fixed effects, even when we focus on identical twins only. Moreover, the coefficients are close in magnitude to those from the OLS regression, which are overall, quite comparable to the findings of previous studies. This is reassuring as it indicates that subjective well-being studies, despite using data from different countries, different scales and populations, applying different estimation techniques, still manage to persistently capture the most important factors that determine happiness.

Table 5.2 Estimates for determinants of emotional well-being

	OLS	FE	FE, MZ
Age	-0.013^{**}		
	[0.005]		
Male	0.099^{***}	0.098^{*}	
	[0.027]	[0.051]	
Married	0.267^{***}	0.304^{***}	0.275^{***}
	[0.029]	[0.043]	[0.067]
Divorced/separated	-0.293^{***}	-0.336^{***}	-0.205
	[0.057]	[0.081]	[0.127]
Education (in years)	0.026^{***}	0.032^{***}	0.059^{***}
	[0.005]	[0.010]	[0.015]
Children	-0.029^{**}	-0.016	-0.027
	[0.014]	[0.020]	[0.034]
Religious	0.044	0.041	-0.064
	[0.027]	[0.048]	[0.072]
Health (1-4)	0.353^{***}	0.322^{***}	0.244^{***}
	[0.014]	[0.019]	[0.031]
Extroverted	0.115^{***}	0.055^{**}	-0.009
	[0.014]	[0.026]	[0.041]
Neurotic	-0.144^{***}	-0.086^{**}	-0.035
	[0.015]	[0.024]	[0.040]
Unemployed	-0.202^{**}	-0.071	0.201
	[0.082]	[0.113]	[0.173]
Low income	-0.216^{***}	-0.189^{***}	-0.233^{*}
	[0.057]	[0.076]	[0.138]
High income	0.010	-0.047	-0.060
	[0.040]	[0.058]	[0.086]
Observations	5524	5524	2326

Notes: FE stands for fixed effects, MZ for monozygtic twins; Standard erros in parantheses, clustered within twin pairs; EWB, health, extroversion and neuroticism are standardized to mean 0, variance 1; * indicates statistically significant at the 10% level, ** at the 5%, and *** at the 1% level

5.5.2 Estimating the effect of the traumatic factors

In Table 5.3 we repeat the estimations so far but also include the variables for traumatic events in childhood and adulthood.[89] In the first column, we have again a linear relationship. Including the controls for traumatic events, we are explaining around 24% of the variance of the EWB compared to 21.6% without them. All of the traumatic events in childhood – physical and sexual abuse and neglect – are statistically significant and negatively associated with the current EWB. If the respondent has been in jail, that is also

[8] We estimated both models with and without extra controls. Those without controls in general explained less of the EWB variability and had higher in absolute value coefficients. We focus on models with full set of controls to increase the predictability of our model. In the FE, in general, the same variables had statistical significance.

[9] We also collapsed the different traumatic events only into a few categories, such as sexually-related traumas, violent traumas and accidental traumas in order to make sure it is not a few unlucky individuals who are driving the results. The estimates were comparable to Table 5.3.

likely to contribute negatively to his/her well-being. From the traumatic events, reporting rape significantly reduces the well-being and witnessing an injury/murder significantly increases it. The latter could simply be a spurious relationship, or could indicate that if someone experienced something traumatic, in which he/she was not directly harmed, this would prompt them to appreciate and value life more.

In the whole sample of twins – given in column (2) – reporting physical or sexual abuse has a negative and statistically significant effect on the reported subjective well-being, as well as ever doing something you could be arrested for, even though you were not. Ever being involved in an accident reduces the EWB with 0.08 sd. Among MZ twins, only being assaulted is statistically significant at the 5% level. If someone reported assault at any point in their life, this reduces their EWB with 0.21 standard deviations. We need to note though that since it varies how long ago the traumatic events took place, the coefficients could reflect different level of adaptation. However, Table 5.3 does not provide convincing an overall evidence that traumatic experiences affect one's long-run well-being once we remove the common family environment and genetics.

5.5.3 Testing the adaptation hypothesis with traumatic events in the past 3 and 1 year

In this section we test the validity of the adaptation hypothesis by focusing on traumatic events that happened in the past 3 years and in the past year, respectively. If it is indeed the case that humans do adapt to the circumstances, then we expect to find a stronger effect of more recent events. The results for the sample of monozygotic twins are given in Table 5.4 below.

Table 5.3 Estimates for determinants of emotional well-being, including traumatic events

	OLS	FE	FE, MZ
Controls	Yes	Yes	Yes
Physical abuse	-0.089^{***}	-0.097^{**}	-0.068
	[0.025]	[0.038]	[0.057]
Sexual abuse	-0.138^{***}	-0.195^{***}	0.026
	[0.046]	[0.067]	[0.114]
Neglect	-0.362^{***}	-0.235	0.259
	[0.141]	[0.188]	[0.306]
Ever arrested	-0.053	-0.033	0.110
	[0.082]	[0.110]	[0.163]
Ever in jail	-0.278^{**}	-0.267	-0.297
	[0.122]	[0.231]	[0.391]
Could be arrested	-0.116^{***}	-0.121^{***}	-0.073
	[0.031]	[0.044]	[0.071]
Was in accident	-0.068^{**}	-0.080^{*}	-0.089
	[0.033]	[0.047]	[0.071]
Was in disaster	-0.002	0.053	-0.025
	[0.034]	[0.051]	[0.080]
Was held captive	0.064	0.098	0.055
	[0.048]	[0.068]	[0.099]
Was raped	-0.167^{***}	-0.137	0.031
	[0.063]	[0.092]	[0.139]
Was assaulted	-0.045	-0.072	-0.212^{**}
	[0.046]	[0.064]	[0.100]
Witnessed injury/murder	0.094^{***}	0.084^{*}	-0.012
	[0.046]	[0.044]	[0.066]
Trauma sb else	-0.080^{*}	-0.058	0.084
	[0.046]	[0.065]	[0.096]
Observations	5524	5524	2326

Note: See Table 5.2

We see that the marital and health status and the years of education continue to be positive and statistically significant. Interesting patterns emerge from the recent adverse experiences as well. If someone has been in an accident, or was assaulted in the past 3 years, this reduces his/her EWB with, all else being equal, with 0.4 and 0.51 standard deviations, respectively. When we focus on the adverse events from the past year, all the above mentioned variables preserve their significance and magnitude, plus now reporting rape significantly reduces the well-being with around 0.75 standard deviations. Comparing our findings in Tables 5.4 and 5.3, we see that overall more recent traumatic events seem to matter more for the subjective well-being, which confirms the overall validity of the adaptation hypothesis.

Table 5.4 Effects of recent traumatic events on the emotional well-being

	FE, MZ, traumas in the past 3 years	FE, MZ, traumas in the past year
Married	0.253^{***}	0.240^{***}
	[0.064]	[0.065]
Divorced	-0.214^{*}	-0.223^{*}
	[0.126]	[0.126]
Education (in years)	0.060^{***}	0.060^{***}
	[0.015]	[0.015]
Health	0.246^{***}	0.244^{***}
	[0.031]	[0.031]
Religion	−0.074	−0.057
	[0.071]	[0.072]
Unemployed	0.193	0.187
	[0.172]	[0.173]
Low income	−0.195	0.198
	[0.135]	[0.136]
High income	−0.058	−0.052
	[0.085]	[0.084]
Physical abuse	−0.068	−0.068
	[0.055]	[0.056]
Sexual abuse	0.034	0.028
	[0.111]	[0.111]
Neglect	0.253	0.231
	[0.286]	[0.279]
Ever arrested	0.130	0.095
	[0.155]	[0.155]
Ever in jail	−0.340	−0.335
	[0.378]	[0.363]
Could be arrested	−0.091	−0.085
	[0.070]	[0.070]
Recent accident	-0.413^{***}	-0.413^{***}
	[0.124]	[0.157]
Recent disaster	−0.050	−0.090
	[0.134]	[0.205]
Recent assault	-0.487^{**}	-0.672^{**}
	[0.191]	[0.298]
Recent rape	−0.024	-0.753^{**}
	[0.284]	[0.360]
Recently witness injury/murder	0.004	−0.004
	[0.113]	[0.154]
Recently held captive	0.452^{*}	0.667^{*}
	[0.237]	[0.372]
Trauma sb else	0.070	0.087
	[0.094]	[0.093]
Observation	2326	2326

Note: See Table 5.2

5.6 Sensitivity checks

In this section we try to account for a few problems that could threaten the internal validity of our estimation approach. First of all, we try to investigate whether the current mood

affects the provided answers. Then, we try to account for a potential reverse causality by using information from the first wave. Finally, we check whether there is a direct effect of some of the co-twin's observable characteristics on the EWB of his/her sibling.

5.6.1 Measurement error

The mood of the respondent during the interview could affect both the provided answers to the EWB questions and the given responses to some of the explanatory variables. First of all, to check to what extent current mood affects the answers provided (and also to account for one of the twins being, in general, more negative than his co-twin), we regressed the difference in the ratings of the emotional well-being on the difference in the answers of the questions that should be answered in the same way, such as whether the respondents were raised by both natural parents, whether the parents used to fight in front of the children, whether either of the parents had problems with alcohol, how often the twins see and contact each other. Some of these questions are neutral, and some are more sensitive, and one can argue that if someone is more negative in general, or was in a bad mood, this could be shown in the answers to some of these questions. The results are displayed in Table 5.5 below. Each row gives the result of a separate regression. Only the difference of how often the twins see each other when regressed on the difference in the EWB ratings is statistically significant, but this is likely to be just a spurious relationship, since a very similar variable – how often the twins contact each other – is not different from zero. So, overall Table 5.5 does not provide compelling evidence of a systematic misreporting because of the current mood or general pessimism/optimism.

Table 5.5 Difference in EWB explained by differences in answers to questions that should have been answered the same way

	△ EWB
△raised by both nat. parents	0.072
	[0.061]
△parents fight in front of children	0.020
	[0.029]
△mom drinking problem	0.012
	[0.061]
△dad drinking problem	0.007
	[0.039]
△see each other	0.115**
	[0.041]
△contact each other	0.010
	[0.032]

Note: standard errors given in brackets []

5.6.2 Reverse causality

Since one might argue that it is happy people who enjoy a higher probability of marriage and better health, we used marital status and health information from a previous wave. The data were gathered 1989-1990. Unfortunately, the first questionnnaire does not contain information on the emotional well-being, but only includes health behavior (and problems) and marital status. We instrumented the marital status at the time of the second wave with the marital status at the time of the first wave. We found that the first stage of the 2SLS is strong, and the marital status in the past is a strong predictor of the current marital status (coefficient of 0.2, statistically significant at the 1% level). However, in the second stage, the marital status is no longer significant (for the whole sample of twins, and the MZ samples). A similar pattern is found for health status. We do not find an effect of marital status and health measured at the first wave on the emotional well-being measured in the second wave but we acknowledge this is not the ideal approach we would like to follow had we more data available.

5.6.3 Comparison to co-twin

One might argue that twins, especially identical twins, are a special case and expect twins

to have a special bond. Studies show that the comparison people make with others is very important in the way they rate their well-being and we check to what extent the well-being and some personal characteristics of the co-twin directly affect his/her sibling's well-being. If one of the twins is doing quite well, enjoys a prosperous job, good health and a happy marriage, his co-twin could either share his happiness, or this could exert a negative effect on his well-being if his own situation is any less favourable. Therefore, we included the co-twin's well-being, marital status, self-reported health, unemployment status and indicators for having either low or high income as right-hand side variables. Note that this will not allow us to perform twin-fixed effects due to perfect collinearity, so we stick to estimating an OLS relationship. Of the included variables, only the marital status and the emotional well-being of the co-twin were significant. If the twin is married, this decreases the EWB of his co-twin with 0.06 of a standard deviation, and one deviation increase in the EWB of a twin increases the EWB of his co-twin with around 0.07 of a standard deviation. So, there is some indication for the effect of the co-twin's situation but since we cannot perform our twin-fixed effects estimation, we refrain from making strong conclusions.

5.6.4 The effect of nature

To distinguish between nature versus nurture, we repeated our main estimates for twins who lived together until they were 18. For pairs who started to live apart from early on, one could argue there is a lower effect of the shared envrionment (lower nurture effect). The majority of our respondents (96%) lived together until they were at least 16, and more than 75% of them lived together until they were at least 18. Repeating our main estimations and the estimations for the effect of traumatic events for twins who lived together until they were 18, we found quite similar results to those in Section 5.5. Therefore, a lower effect of shared environment would not be a problem for our results.

5.7 Discussion and conclusions

Despite the increased economic interest in well-being in the past few decades, there are still quite a few pressing issues in the literature. Establishing causality is difficult and so far, the best attempts in the literature use longitudinal data. Whereas using panel data reduces bias stemming from individual heterogeneity, with such approach we cannot account for time-invariant characteristics of the respondents – a limitation we can overcome with twin data. The biggest advantage of the twin-fixed effects is that we purge our estimates of unobserved familial and genetic similarities, reducing the omitted variables bias. Furthermore, some studies argue that happiness is a genetically determined personality trait. With data on identical twins, such a hypothesis – even if true – would not pose a problem for our results. Of course, this approach comes with its limitations, acknowledged and extensively discussed in the chapter.

Our findings are consistent with the stylized facts in the literature. We find that marital status, self-reported health, years of education and low income have a significant effect on self-reported emotional well-being. Moreover, the magnitude of our coefficients is rather similar to the most prominent studies in the literature. This is a good signal of the usefulness of well-being as a valuable concept with potential important implications. We also confirm the validity of the adaptation hypothesis, which postulates that humans can adjust to negative shocks. We find the strongest effect from traumatic events that happened in the past year, still a strong effect from some adverse experiences from the past three years, and overall, a dissipating effect over time.

As we noted, we were limited in using only a single question to measure the well-being in our sample. We believe that a more extensive analysis using twins could be helpful in answering some pressing questions in the well-being literature. In the future, more complex measures of well-being (focusing on a number of aspects of one's life), questions about one's satisfaction with life, as well as inquiries about co-twin's well-being would be welcome. In this way, we could better distinguish the difference in well-being measures as

well as gain more understanding about the nature-nurture relationship.

Our results raise a number of policy-relevant implications. First of all, we cannot ignore the important impact of health and education on one's well-being. Therefore, promoting education and healthy behavior is likely to generate returns in terms of well-being, among other things. Eradicating poverty and promoting relationship skills is likely to contribute to society's well-being as well. Finally, victims of different traumatic experiences should be assisted in order to recover more quickly from their ordeals. All of these policies could lead to large potential returns and contribute to a happier society.

Appendix

Table A1. Summary statistics, all covariates

	All	DZ twins	MZ twins	Males	Females
Emotional well-being	3.16	3.15	3.19	3.21	3.12
[1,2,3,4]	[0.7]	[0.71]	[0.68]	[0.69]	[0.71]
Age	29.9	29.9	29.9	29.9	29.9
	[2.47]	[2.47]	[2.46]	[2.45]	[2.49]
Gender (1 if male)	0.45	0.46	0.41	1	0
	[0.50]	[0.50]	[0.49]		
Marital status (1 if	0.64	0.65	0.65	0.61	0.67
married/cohabiting)	[0.48]	[0.48]	[0.48]	[0.49]	[0.47]
Divorced/separated	0.07	0.07	0.07	0.05	0.08
	[0.25]	[0.25]	[0.25]	[0.23]	[0.26]
Married/cohabitating	0.23	0.24	0.24	0.16	0.28
at time t1	[0.42]	[0.43]	[0.43]	[0.37]	[0.45]
Education (in years)	12.1	12.19	12.27	12.07	12.21
	[2.5]	[2.46]	[2.48]	[2.47]	[2.45]
Education co-twin	11.9	11.95	12.10	11.9	11.9
(in years)	[2.38]	[2.37]	[2.39]	[2.4]	[2.37]
Health status	3.1	3.09	3.15	3.12	3.11
	[0.69]	[0.69]	[0.67]	[0.69]	[0.69]
Any health problems	0.24	0.24	0.27	0.21	0.26
at t1	[0.43]	[0.42]	[0.45]	[0.41]	[0.44]
Religious (1 if yes)	0.69	0.68	0.71	0.64	0.73
	[0.46]	[0.47]	[0.45]	[0.48]	[0.45]
Extroverted (0-12	4.10	4.15	4.52	3.41	4.67
scale)	[4.63]	[4.63]	[4.66]	[4.43]	[4.71]
Neuroticism (0-12	2.66	2.69	2.92	1.94	3.25
scale)	[3.43]	[3.45]	[3.46]	[2.99]	[3.64]
Smoker at t1	0.22	0.24	0.21	0.22	0.23
	[0.42]	[0.42]	[0.41]	[0.41]	[0.42]
Regular drinking at	0.37	0.39	0.39	0.42	0.33
t1	[0.48]	[0.49]	[0.49]	[0.49]	[0.47]
Unemployed	0.04	0.04	0.04	0.05	0.03
	[0.19]	[0.19]	[0.19]	[0.22]	[0.17]
Income in lowest	0.07	0.07	0.07	0.05	0.09
quartile	[0.26]	[0.26]	[0.25]	[0.22]	[0.28]
Income in highest	0.11	0.10	0.12	0.12	0.10
quartile	[0.31]	[0.30]	[0.33]	[0.33]	[0.31]
Physical abuse	0.34	0.34	0.34	0.47	0.26
	[0.49]	[0.49]	[0.49]	[0.50]	[0.48]
Sexual abuse	0.12	0.12	0.11	0.06	0.15
	[0.31]	[0.32]	[0.31]	[0.23]	[0.36]
Neglect	0.02	0.01	0.01	0.01	0.02
	[0.13]	[0.11]	[0.12]	[0.11]	[0.14]
Ever arrested	0.04	0.03	0.03	0.05	0.01
	[0.17]	[0.17]	[0.16]	[0.22]	[0.11]
Ever been in jail	0.01	0.01	0.009	0.02	0.003
	[0.11]	[0.11]	[0.09]	[0.14]	[0.06]
Ever done something	0.25	0.27	0.22	0.37	0.15
that could be	[0.43]	[0.44]	[0.41]	[0.48]	[0.36]
arrested for					

Has been in an accident	0.19 [0.39]	0.20 [0.40]	0.18 [0.39]	0.27 [0.44]	0.13 [0.33]
Has experienced natural disaster	0.14 [0.34]	0.15 [0.35]	0.13 [0.33]	0.17 [0.37]	0.11 [0.32]
Has been assaulted	0.10 [0.30]	0.10 [0.30]	0.09 [0.29]	0.15 [0.35]	0.06 [0.24]
Has been raped	0.05 [0.22]	0.05 [0.22]	0.06 [0.23]	0.02 [0.12]	0.08 [0.28]
Has been held captive	0.08 [0.27]	0.07 [0.25]	0.08 [0.28]	0.11 [0.32]	0.05 [0.22]
Has witnessed a serious injury (or murder)	0.23 [0.42]	0.24 [0.43]	0.22 [0.41]	0.32 [0.47]	0.16 [0.37]
Serious trauma to somebody else	0.09 [0.30]	0.08 [0.27]	0.09 [0.28]	0.08 [0.28]	0.09 [0.28]
Observations	6265	3198	2332	2803	3462

Note: standard deviations are given in brackets []

Summary

Negative experiences can have long-run harmful effects. Numerous policies aim to improve the situation of people who have been through something traumatic; governments spend billions to fight and prevent crime, and improve the quality of education and well-being of their citizens. But without precise quantitative evidence, the efficiency of the different policies is questionable. Aiming to minimize social costs, economists have been focusing on studying causal relationships. The effectiveness of different policy measures has been analyzed but many pressing questions remain largely neglected. In this dissertation, we have paid attention to a few such problems.

The first two chapters analyze a similar problem. In Chapter 2, we investigate the effect of physical and sexual child maltreatment on several types of illegal and problematic behavior. The estimated effects suggest an increase of illegal and problematic behavior between fifty and one hundred percent. Our findings are consistent with the so-called 'cycle of violence' hypothesis. Given our findings and the importance of the problem, policies should be designed to not only alleviate the long-run impact on victims of child maltreatment but to also prevent the problem before occurring. We showed that child maltreatment is more frequent among families who have overall more problems. Therefore, extra assistance and help to children and parents from such families could be a good initial preventive measure.

In Chapter 3 we investigate a more general formulation of this problem. We analyze the link between victimization and offending by using a representative sample from the Netherlands and employing different estimation approaches. We fail to find convincing evidence for a strong effect of past victimization on the current offending, both in the

short and the long run. One might think that our findings in Chapters 2 and 3 contradict each other. However, we need to keep in mind that we discuss very different negative "treatments" in them. In Chapter 2 the maltreatment takes place during childhood, thus it could occur over a longer period of time and affect an individual during their most vulnerable age, whereas in Chapter 3 the acts of victimization are usually one-time events, which happened when the individuals were adults and are not always as traumatic as child maltreatment (one can imagine that having your car broken into or a bike stolen could have less of an impact than experiencing child maltreatment).

Chapter 4 is slightly different than the others. Namely, in it we investigate the effect of age differences within grades among students in primary education. We find a large gap of about 0.9 standard deviations in language and math test scores for pupils in grade 2, which decreases to about 0.3 standard deviations by grade 8, and remains statistically significant. Interestingly, for students from low socio-economic background, the gap in grade 8 is 74% larger than for eight-graders who do not have a disadvantaged background. While we do not per se discuss how education can smooth and alleviate the negative shocks we explore in Chapters 2, 3, and 5, education might have an impact on the way people deal with such negative experiences as it equips students with skills and knowledge and can foster resilience towards negative shocks.

In Chapter 5, we again use data on twins to explore emotional well-being and its determinants. We find that in the monozygotic twin fixed effects estimations, the marital status, health, years of education, and having low income preserve their significance, thus confirming the most pronounced stylized facts in the happiness literature. Moreover, we find strong negative effect of more recent traumatic events, such as being assaulted, being raped or taking part in an accident; thus, we confirm the validity of adaptation. These findings confirm the need for policies and assistance to victims of different traumatic experiences.

Nederlandse Samenvatting (Summary in Dutch)

Negatieve ervaringen kunnen lange termijn schadelijke effecten veroorzaken voor degene die ze ervaren. Veel beleidsmaatregelen zijn gericht om de situatie van mensen die iets traumatisch hebben ervaringen te verbeteren. Overheden spenderen miljarden aan het bestrijden en voorkomen van criminaliteit, verbeteren van de kwaliteit van educatie en welzijn van haar burgers. Maar zonder nauwkeurig kwantitatief bewijsmateriaal is de efficiëntie van de verschillende beleidsterreinen twijfelachtig. Om sociale kosten te minimaliseren focussen economen zich op het bestuderen van causale relaties. De effectiviteit van verschillende maatregelen is geanalyseerd, maar veel prangende vragen blijven grotendeels onbeantwoord . In deze proefschrift, hebben we aandacht besteed aan een paar van zulke problemen.

De eerste twee hoofdstukken is er een soortgelijk probleem geanalyseerd. In hoofdstuk 2 onderzoeken we het effect van fysieke en seksuele kindermishandeling op verschillende soorten crimineel probleemgedrag . Door het gebruik van variatie binnen tweelingen zijn we in staat om verstorende factoren uit eerdere studies te mitigeren. Met behulp van "Ordinary Least Squares" en "twin fixed effects" concluderen we dat kindermishandeling een groot effect heeft op crimineel probleemgedrag . Zoals: drugsgebruik, gedragsstoornissen en criminaliteit. De geschatte effecten wijzen op een toename van crimineel probleemgedrag tussen de vijftig en honderd procent. Onze bevindingen zijn consistent met de zogenaamde "cycle of violence" hypothese.

In Hoofdstuk 3 onderzoeken we een meer algemene formulering van dit probleem. We analyseren de relatie tussen slachtofferschap en criminaliteit door het gebruik van een

representatieve steekproef uit Nederland met gebruikmaking van verschillende schatting methoden. We beginnen met "Ordinary Least Squares". Indien van toepassing, gericht op verschillende samples van onze data met een aantoonbaar lager selectie vooringenomenheid. Daarna gebruiken we een strategie voorgesteld door Altonji et al. (2005), die de selectie van de waarneembare factoren gebruikt als leidraad voor de selectie van niet-waarneembare factoren. Bij het gebruik van deze methoden vinden we geen sterk bewijs dat het verleden van een slachtoffer effect heeft op huidige crimineel gedrag, zowel op de korte als de lange termijn.

In hoofdstuk 4 onderzoeken we het effect van leeftijd op verschillen klassen in testscores van leerlingen in het basisonderwijs. Specifiek onderzoeken we of deze relatieve leeftijdseffecten veranderen tijdens het basisonderwijs en welke schoolfactoren aan deze verandering bijdragen. We vonden een groot hiaat van ongeveer 0,9 standaarddeviaties in taal en wiskunde test scores voor leerlingen in leerjaar 2. Deze neemt af tot ongeveer 0,3 standaarddeviaties voor groep 8, en blijft statistisch significant. We vonden een hiaat tussen de oudste en de jongste leerlingen op een "high-stakes" test. Dit resulteerde in dat jongste leerlingen minder vaak doorstromen naar secundair onderwijs . De grootste bijdrage in de daling van de jongste-oudste prestaties gap is het looptijdseffect. Bijv. jongere studenten kunnen hun oudere collega's inhalen. Het sturen van leerlingen naar het speciaal onderwijs en logopedie is ook belangrijk. Tot slot, voor studenten uit een laag sociaal-economische achtergrond is het gat in groep 8 74% groter dan voor achtste groepers van reguliere achtergrond.

In Hoofdstuk 5 gebruiken we weer tweelingdata om emotioneel welzijn en de determinanten te verkennen. We hebben drie dingen gerealiseerd. In de eerste plaats, met behulp van twin fixed effecten, testen we de robuustheid van eerdere studies in welzijn literatuur. Wij vinden dat in de identical twin fixed effecten schattingen: de burgerlijke staat, gezondheid, het onderwijs, en een laag inkomen hun betekenis behouden . Waardoor de meest uitgesproken feiten in literatuur over geluk worden bevestigd. Ten tweede in het gebruik van informatie over traumatische gebeurtenissen, testen we de validiteit van de

hypothese. Volgens welke de mens zich kan aanpassen aan zowel positieve als negatieve schokken en terug te keren naar een een gewenste mate van tevredenheid met het leven. We zien een sterk negatief effect van meer recente traumatische gebeurtenissen. Zoals aanranding, verkrachting of deel te nemen aan een ongeval. Hiermee bevestigen wij de geldigheid van de aanpassing hypothese. Tenslotte tonen wij dat genetische disposities belangrijk zijn voor de intra-pair variantie van het emotionele welbevinden.

Bibliography

Agnew, Robert. 1985. A revised strain theory on delinquency. *Social Forces*, **64**, 151–167. 44

Agnew, Robert. 2001. Building on the foundation of general strain theory: specifying the types of strain most likely to lead to crime and delinquency. *Journal of Research in Crime and Delinquency*, **38**, 319–361. 44, 46, 47

Agnew, Robert. 2002. Experienced, vicarious and anticipated strain: an exploratory study on physical victimization and delinquency. *Justice Quarterly*, **19**, 603–632.

Agnew, Robert, Brezina, Timothy, Wright, John Paul, & Cullen, Francis T. 2002. Strain, Personality traits, and delinquency: extended general strain theory. *Criminology*, **40**(1).

Almond, Douglas. 2006. Is the 1918 influenza pandemic over? Long-term effects of in-utero influenza exposure in the post-1940 US population. *Journal of Political Economy*, **114**, 672–712. 76

Altonji, Joseph G., Elder, Todd E., & Taber, Christopher R. 2005. Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools. *Journal of Political Economy*, **113**(1), 151–184. 44

Altonji, Joseph G., Elder, Todd E., & Taber, Christopher R. 2008. Using Selection on Observed Variables to Assess Bias from Unobservables when Evaluating Swan-Ganz Catheterization. *American Economic Review: Papers and Proceedings*, **98**(2), 345–350.

Andrews, Frank M., & Whitey, Stephen B. 1976. *Social Indicators of Well-Being: America's Perception of Life Quality*. Plenum, New York.

Angeles, Luis. 2010. Adaptation and Anticipation Effects to Life Events in the United Kingdom. *Discussion Paper, Department of Economics, University of Glasgow, Glasgow.* 109

Ashenfelter, Orley, & Krueger, Alan. 1994. Estimates of the economic return to schooling from a new sample of twins. *The American Economic Review*, **84**(5), 1157–1173. 37, 113

Baetschmann, Gregory, Staub, Kevin E., & Winkelmann, Rainer. 2011. Consistent estimation of fixed effects ordered logit. *IZA DP No*, **5443.** 112

Barschak, Erna. 1951. A study of happiness and unhappiness in the childhood and adolescence of girls in different cultures. *Journal of Psychology*, **32**, 173–215. 109

Bartels, Meike, & Boomsma, Dorret I. 2009. Born to be Happy? The Etiology of Subjectiv Well-Being. *Behavior Genetics*, **39**(605). 104

Becchetti, Leonardo, & Pelloni, Alessandra. 2013. What are we learning from the life satisfaction literature? *International Review of Economics*, **60**, 113–155. 105, 106

Becker, Gary S. 1968. Crime and punishment: an economic approach. *Journal of Political Economy*, **76**(2), 169–217. 11

Bedard, Kelly, & Dhuey, Elizabeth. 2006. The Persistence of Early Childhood Maturity: International Evidence of Long-run Age Effects. *The Quarterly Journal of Economics*, **121**(4), 1437–1472. 3, 70, 76

Benda, Brent B., & Corwyn, Robert F. 2002. The Effect of abuse in childhood and in adolescence on violence among adolescents. *Youth Society*, **33**(39), 339–36. 11

Bickel, Donna DiPrima, Zigmond, Naomi, & Strayhorn, Joseph. 1991. Chronological Age at Entrance to First Grade: Effects on Elementary School Success. *Early Childhood Research Quarterly*, **6**, 105–117. 3, 70

Bjornskov, Christian, Dreher, Axel, & Fischer, Justina A. V. 2008. Cross-country determinants of life satisfaction: Exploring different determinants across groups in society. *Social Choice and Welfare*, **30**, 119–173. 105

Black, Sandra E., Devereux, Paul J., & Salvanes, Kjell G. 2011. Too Young To Leave The Nest? The Effects of School Starting Age. *The Review of Economics and Statistics*, **93**(2), 455–467. 70, 73

Bohnke, Petra, & Kohler, Ulrich. 2007. Determinants of Happiness: Life Satisfaction in Europe. *WSI Mitteillungen*, **7**(373). 106

Bound, John, & Solon, Gary. 1999. Double Trouble: On the Value of Twins-Based Estimation of the Return to Schooling. *Economics of Education Review*, **18**(2), 169–82. 25, 26

Brooks, D N, & McKinlay, W. 1983. Personality and Behavioral Change after Severe Blunt Head Injury - A Relative's View. *Journal of Neurology, Neurosurgery and Psychiatry*, **46**(336). 106

Brorsson, B, Ifver, J., & Hays, R.D. 1993. The Swedish Health-Related Quality of Life Survey. *Quality of Life Research*, **2**(33). 109

Bruni, L, & Stanca, L. 2006. Watching alone: Relational goods, television and happiness. *Journal of Economic Behavior and Organization*, **65**, 506–528. 106

Buckels, Kasey, & Hungerman, Daniel M. 2013. Season of Birth and Later Outcomes: Old Questions, New Answers. *The Review of Economics and Statistics*, **95**(3), 711–724. 76

Cameron, Colin A., & Trivedi, Pravin K. 2010. *Microeconometrics Using Stata, Revised Edition*. Stata press.

Caspi, Avshalom, & Moffitt, Terrie E. 1995. The continuity of maladaptive behaviour: From description to understanding in the study of antisocial behaviour. *Developmental Physcopathology*, **2: Risk, disorder and adaptation**. 11

CBS. 2009 (september). *Many suspects are also victims.*

Chen, Xiaojin. 2009. The link between juvenile offending and victimization: the influence of risky lifestyles, social bonding, and individual characteristics. *Youth Violence and Juvenile Justice*, **7**. 46

Clark, Andrew. 1997. Job satisfaction and gender: Why are women so happy at work? *Labor Economics*, **4**(4), 341–72. 105

Clark, Andrew, & Oswald, Andrew. 1994. Unhappiness and unemployment. *The Economic Journal*, **104**, 648–659. 107, 115

Clark, Andrew, & Oswald, Andrew. 1996. Satisfaction and comparison income. *Journal of Public Economics*, **61**, 359–381.

Clark, Andrew E., Diener, Ed, Georgellis, Yannis, & Lucas, Richard E. 2008. Lags and leads in life satisfaction: A test of the baseline hypothesis. *The Economic Journal*, **118**, 222–243. 117

Cohen, Lawrence E., & James R. Kluegel, Kenneth C. Lan. 1981. Social Inequality and Predatory Criminal Victimization: An Exposition and Test of a Formal Theory. *American Sociological Review*, **46**(5), 505–524.

Coombs, Robert H. 1991. Marital status and personal well-being: A literature review. *Family Relations*, **40**, 97–102. 106

Cornwell, Christopher, Mustard, David B., & Parys, Jessica Van. 2013. Noncognitive Skills and the Gender Disparities in Test Scores and Teacher Assessment. *The Journal of Human Resources*, **48**(1), 236–264. 74

Costa, Paul T., & McCrae, Robert R. 1980. Influence of extroversion and neuroticism on subjective well-being: Happy and unhappy people. *Journal of Personality and Social Psychology*, **38**, 668–678. 109, 115

Crawford, Claire, Dearden, Lorraine, & Greaves, Ellen. 2014. The drivers of month-of-birth differences in children's cognitive and non-cognitive skills. *Journal of the Royal Statistical Society, Series A*, 829–860. 73

Cuhna, Fabio, & Heckman, James J. 2007. The Technology of Skill Formation. *AEA Papers and Proceedings*, **92**(2), 31–47.

Cuhna, Fabio, Heckman, James J, Lochner, Lance, & Masterov, Dimitry V. *Interpreting the evidence of life cycle skill formation.*

Cullen, Francis T., Unnever, James D., Hartman, Jennifer L., Turner, Michael G., & Agnew, Robert. 2008. Gender, bullying, victimization, and juvenile delinquency: a test of general strain theory. *Victims and Offenders*, **3**, 346–364.

Currie, Janet, & Tekin, Erdal. 2012. Understanding the cycle: Childhood maltreatment and future crime. *Journal of Human Resources*, **47**(2.). 2, 7, 11, 16, 46

Currie, Jannet, & Stabile, Mark. 2003. Socioeconomic Status and Health: Why is the Relationship Stronger for Older Children? *American Economic Review*, **93**(5), 1813–1823. 10

Daigle, Leah E., Cullen, Francis T., & Wright, John Paul. 2007. Gender difference in the predictors of juvenile delinquency: assessing the generality/specificity debate. *Youth Violence and Juvenile Justice*, **5**, 254–286.

Dembo, Richard, Dertke, Max, LaVoie, Lawrence, Borders, Scott, Washburn, Mark, & Schmeidler, James. 1987. Physical abuse, sexual victimization and illicit drug use: a structural analysis among high risk adolescents. *Journal of Adolescence*, 13–33. 12

Dhuey, Elizabeth, & Lipscomb, Stephen. 2010. Disabled or young? Relative age and special education diagnoses in schools. *Economics of Education Review*, **29**, 857–872. 74

Diener, Ed. 1994. Assessing subjective well-being: Progress and opportunities. *Social Indicators Research*, **31**, 103–157.

Diener, Ed, Sandvik, Ed, Pavot, William, & Gallagher, Dennis. 1991. Response artifacts in the measurement of subjective well-being. *Social Indicators Research*, **24**, 35–56. 115

Diener, Ed, Suh, Eunkook M., Lucas, Richard E., & Smith, Heidi L. 1999. Subjective well-being: Three decades of progress. *Psychological Bulletin*, **125**(2), 276–302. 108, 112, 117

Dinwiddie, Stephen H., Heath, A. C., Dunne, M. P., Bucholtz, Kathleen K., Madden, Pamela A. F., Slutske, W. S., Bierut, L. J., Statham, Dixie B., & Martin, Nicholas G. 2000. Early sexual abuse and lifetime psychopathology: a co-twin- control study. *Psychological Medicine*, **30**, 41–52. 6, 13

Dittmann, Joerg, & Goebel, Jan. 2010. Your House, Your Car, Your Education: The Socioeconomic Situation of the Neighborhood and its Impact on Life Satisfaction in Germany. *Social Indicators Research*, **96**(497). 105, 106

Dodge, Kenneth A., & James E. Bates, Gregory S. Pettit. 1990. Mechanisms in the Cycle of Violence. *Science*, **250**, 1678–1683. 11

Dolan, Paul, Peasgood, Tessa, & White, Mathew. 2008. Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being. *Journal of Economic Psychology*, **29**, 94–122. 121

Downey, Douglas B., & Pribesch, Shana. 2004. When Race Matters: Teachers' Evaluations of Students' Classroom Behavior. *Sociology of Education*, **77**, 267–282. 74

Dube, Shanta R., Felitti, Vincent J., Dong, Maxia, Chapman, Daniel P., Giles, Wayne H., & Anda, Robert F. 2003. Childhood Abuse, Neglect, and Household Dysfunction and the Risk of Illicit Drug Use: The Adverse Childhood Experiences Study. *Pediatrics*, **111**. 12

Dujardin, Claire, & Goffette-Nagot, Florence. 2009. Neighborhood effects on unemployment? A test a la Altonji. *GATE Working papers 09-24*. 55

Dunne, Michael P., Purdie, David M., Cook, Michelle D., Boyle, Frances M., & Najman, Jake M. 2003. Is child sexual abuse declining? Evidence from a population-based survey of men and women in Australia. *Child Abuse and Neglect*, **27**(2), 141–152. 16, 117

Easterlin, Richard A. 1974. Does economic growth improve the human lot? Some empirical evidence. *Nations and Households in Economic Growth. Essays in Honor of Moses Abramovitz*, **89**, 89–125. 102

Easterlin, Richard A. 2001. Income and happiness: Towards a unified theory. *The Economic Journal*, **11**(472), 465–484.

Easterlin, Richard A. 2003. Explaining happiness. *Contribution to special series of Inaugural Articles by members of the National Academy of Sciences*, **100**(19), 11176–11183. 108

Eder, Donna. 1981. Ability Grouping and a Self-Fulfilling Prophecy: A Micro-Analysis of Teacher-Student Interaction. *Sociology of Education*, **54**(3), 151–162. 74

Eitle, David, & Turner, R. Jay. 2002. Exposure to Community Violence and Young Adult Crime: The Effects of Witnessing Violence, Traumatic Victimization, and Other Stressful Life Events. *Journal of Research in Crime and Delinquency*, **39**(2), 214–237.

Elder, Glen H. 1974. *Children of the Great Depression*. University of Chicago Press. 109

Elder, Todd E., & Lubotsky, Darren H. 2009. Kindergarten Entrance Age and Children's Achievement.Impacts of State Policies, Family Background, and Peers. *The Journal of Human Resources*, **44**(3), 641–683. 74

English, Diana J., Widom, Cathy S., & Brandford, Carol. 2002. *Childhood Victimization and Delinquency, Adult Criminality, and Violent Criminal behaviour: A Replication and Extension, Final Report*. Presented to the US National Institute of Justice. 13

Entorf, Horst. 2014. Criminal Victims, Victimized Criminals, or Both? A Deeper Look at the Victim-Offender Overlap. *IZA discussion papers*, **7686**. 47

Evans, William N., Morrill, Melinda S., & Parente, Stephan T. 2010. Measuring inappropriate medical diagnosis and treatment in survey data: The case of ADHD among school-aged children. *Journal of Health Economics*, **29**, 657–673. 74

Fagan, Jeffrey, Piper, Elizabeth S., & Cheng, Yu-Teh. 1987. Contributions of Victimization to Delinquency in Inner Cities. *The Journal of Criminal Law and Criminology*, **78**(3.).

Fajnzylber, Pablo, Lederman, Daniel, & Loayza, Norman. 2000. Crime and Victimization: An Economic Perspective. *Economia*, 219–302.

Ferreri-Carbonell, Ada, & Frijters, Paul. 2002. How important is methodology for the estimates of the determinants of happiness? *Tinbergen Institute Discussion Paper 024/3*. 105, 120

Finkelhor, David, & Browne, Angela. 1985. The Traumatic Impact of Child Sexual Abuse: A Conceptualization. *American Journal of Orthopsychiatry*, **55**. 10

Forde, David R., & Kennedy, Leslie W. 2006. Risky lifestyles, routine activities and the General Theory of Crime. *Justice Quarterly*, **14**(2), 265–294.

Fredricksson, Peter, & Ockert, Bjorn. 2013. Life-cycle effects of age at school start. *The Economic Journal*, **121**, 1437–1472. 3, 70

Frey, Bruno, & Stutzer, Alois. 2002a. The Economics of happiness. *World Economics*, **3**(1.). 107

Frey, Bruno, & Stutzer, Alois. 2002b. What can economists learn from happiness research? *Journal of Economic Literature*, **XL**. 103, 107, 108

Frey, Bruno S., Bentz, Mathias, & Stutzer, Alois. 2004. Introducing procedural utility: Not only What, but also How matters. *Journal of Institutional and Theoretical Economics*, **160**, 377–401.

Gamble, Amelie, & Garling, Tommy. 2012. The relationship between life satisfaction, happiness, and current mood. *Journal of Happiness Studies*, **13**, 31–45.

Garbarino, James, & Gilliam, Gwen. 1980. *Understanding Abusive Families.* Lexington Mass: Lexington Books. 12

Gil, David G. 1970. *Violence Against Children: Physical child Abuse in the United States.* Harvard University Press. 6

Gluckman, Peter D., Hanson, Mark A., Cooper, Cyrus, & Thornburg, Kent L. 2008. Effects of In-Utero and Early Life Conditions on Adult Health and Disease. *The New England Journal of Medicine*, **359**, 61–73. 10

Gortmaker, Stephen L., Kagan, Jerome, Caspi, Avshalom, & Silva, Phil A. 1997. Daylength during pregnancy and shyness of children: Results from northern and southern hemispheres. *Developmental psychobiology*, **31**, 107–114. 76

Gove, Walter R., Style, Carolyn B., & Hughes, Michael. 1990. The effects of marriage on the well-being of adults: A theoretical analysis. *Journal of Family Issues*, **11**, 4–35. 106

Graham, Carol. 2005. Insights on Development from the Economics of Happiness. *The World Bank Research Observer*, **20**(2), 201–231. 105, 120

Graham, Carol, Eggers, Andrew, & Sukhtankar, Sandip. 2004. Does happiness pay? An exploration based on panel data from Russia. *Journal of Economic Behavior and Organization*, **55**, 319–342. 120

Griliches, Zvi. 1979. Sibling models in data in economics: Beginnings of a survey. *The Journal of Political Economy*, **87**. 25, 113

Hagerty, Michael R., & Veenhoven, Ruut. 2003. Wealth and happiness revisited.Growing wealth of nations does go with greater happiness. *Social Indicators Research*, **64**, 1–27. 106, 108

Halman, Loek. 1987. *Traditie, Secularisatie en Individualisering.* Tilburg University Press. 107

Hay, Carter, & Evans, Michael M. 2006. Violent victimization and involvement in delinquency: examining predictions from general strain theory. *Journal of Criminal Justice*, **34**, 261–271.

Headey, Bruce, & Wooden, Mark. 2004. The Effects of Wealth and Income on Subjective Well-Being and Ill-Being. *Economic Record*, **80**(24).

Headey, Bruce, Kelley, Jonathan, & Wearing, Alex. 1993. Dimensions of Mental Health: Life Satisfaction, Positive Affect, Anxiety and Depression. *Social Indicators Research*, **29**, 63–82. 106, 109

Headey, Bruce, Muffels, Ruud, & Wooden, Mark. 2008. Money does not Buy Happiness... Or does it? A Reconsideration Based on the Combined Effects of Wealth, Income and Consumption. *Social Indicators Research*, **87**(65). 106

Helliwell, John F., & Putnam, Robert D. 2004. The social context of well-being. *Philosophical Transactions of the Royal Society, Biological Sciences*, **359**, 1435–1446. 106, 120

Hindelang, Michael J., Gottfresdson, Michael R., & Garofalo, James. 1978. *Victims of Personal Crime:An Empirical Foundation for a Theory of Personal Victimization*. Cambridge, MA. 43

Hoogerheide, Lennart, Kleibergen, Frank, & van Dijk, Herman K. 2007. Natural conjugate priors for the instrumental variables regression applied to the Angrist-Krueger data. *Journal of Econometrics*, **138**, 63–103. 76

Hunter, Rosemary S., & Kilstorm, Ned. 1979. Breaking the cycle in abusive families. *American Journal of Psychiatry*, **136**. 6

Jacobs, Bruce A., & Wright, Richard. 2010. Bounded rationality, retaliation, and the spread of urban violence. *Journal of Interpersonal Violence*, **25**(10), 1739–1766. 47, 48

Jacques, Scott, & Wright, Richard. 2008. The victimization-termination link. *Criminology*, **46**, 1009–1038.

Jaffee, Sara R., Caspi, Avshalom, Mott, Terrie E., Polo-Tomas, Monica, Price, Tom, & Taylor, Alan. 2004. The limits of child effects: Evidence for genetically mediated child effects on corporal punishment but not on physical maltreatment. *Development Psychology*, **40**(6), 1047–58. 17, 119

Jennings, Wesley G., Higgins, George E., Tewksbury, Richard, Gover, Angela R., & Piquero, Alex R. 2010. A Longitudinal Assessment of the Victim-Offender Overlap. *Journal of Interpersonal Violence*, **25**(12), 2147–2174. 42, 43

Jensen, Gary F., & Brownfield, David M. 1986. Gender, Lifestyles, and Victimization: Beyond Routine Activity Theory. *Violence and Victims*, **1**, 85–99. 43

Kahneman, Daniel. 1999. *Experienced utility and objective happiness: A moment-based approach*. Cambridge University Press and the Russell Sage Foundation, New York.

Kahneman, Daniel, & Deaton, Angus. 2010. High income improves evaluation of life but not emotional well-being. *Psychological and Cognitive Sciences*, **103**(38), 16489–16493. 102, 103, 119, 121

Kahneman, Daniel, & Tversky, Amos. 1979. Prospect theory: An analysis of decision under risk. *Econometrica*, **47**, 263–292.

Kainulainen, Sven. 1998. Life Events and Satisfaction with Life in Different Social Classes, Summary. *Kuopio University Publications, Finland*. 109

Kaplan, Howard B. 1980. *Deviant behaviour in Defense of Self*. New York: Academic Press. 12

Kempe, C Henry, Silverman, Frederic N., Steele, Brandt F., Droegemueller, William, & Silver, Henry K. 1962. The Battered Child Syndrome. *The Journal of the American Medical Association*, **181**(1), 17–24. 6

Kessler, Ronald C, & Essex, Marilyn. 1982. Marital status and depression: The importance of coping resources. *Social Forces*, **61**, 484–507. 106

Kohler, H P, Behrman, J, & Skythe, A. 2005. Partner+Children=Happiness? An Assessment of Fertility and Partnerships on Subjective Well-Being. *University of Pennsylvania, Department of Economics, Philadelphia*. 118, 120

Kozcan, Z. 2013. *Does integration increase life satisfaction?* Cambridge working papersin Economics No. 1314. 105

Krause, Annabelle. 2013. Don't worry, be happy? Happiness and reemployment. *Journal of Economic Behavior and Organization*, **36**, 1–20. 105, 107

Lauritsen, Janet, & Laub, John. 2007. *Understanding the link between victimization and offending: New reflections on an old idea.* NCJ-220695. 42

Loeber, R. 1982. The stability of antisocial and delinquent child behaviour: a review. *Child Development*, **53**, 1431–1446. 11, 20

Lucas, Richard. 2008. Long-Term Disability is Associated with Lasting Changes in Subjective Well-being Evidence from Two Nationally Representative Longitudinal Studies. *Journal of Personality and Social Psychology*, **92**(717). 115

Lucas, Richard E., Diener, Ed, & Suh, Eunkook M. 1996. Discriminant validity of well-being measures. *Journal of Personality and Social Psychology*, **71**(3), 616–628.

Lucas, Richard E, Diener, Ed, Alexander Grob, Eunkook M Suh, & Shao, Liang. 1998. *Cross-cultural evidence for the fundamental features of extroversion: The case against sociability.* Manuscript submitted for publication, University of Illinois at Urbana-Champaigne.

Lykken, D, & Tellegen, A. 1996. Happiness is a stochastic phenomenon. *Psychological Science*, **7**(3.). 103

Lynskey, Michael, Heath, Andrew, Nelson, Elliot, Bucholtz, Kathleen, Madden, Pamela, Slutske, Wendy, Statham, Dixie, & Martin, Nicholas. 2002. Genetic and environmental contributions to cannabis dependence in a national young adult twin sample. **32**, 195–207. 18

Lynskey, Michael, Heath, Andrew, Bucholtz, Kathleen, Slutske, Wendy, Madden, Pamela, Nelson, Elliot, Statham, Dixie, & Martin, Nicholas. 2003. Escalation of drug use in early-onset cannabis users vs cotwin controls. *Journal of the American Medicine Association*, **289**, 427–433. 18

Mamun, Abdullah, Lawlor, Debbie, O'Callaghan, Michael, Bor, William, Williams, Gail, & Najman, Jake. 2007. Does childhood sexual abuse predict young adult's BMI? A birth cohort study. *Obesity*, **8**, 2103–2110. 16, 117

Mayhew, Pat. 2003. Counting the Costs of Crime in Australia. *Trends and issues in crime and criminal justice*, **247**. 7

Mazza, Danielle, & Dennerstein, Lorraine. 2001. The physical, sexual and emotional violence history of middle-aged women: A community-based prevalence study. *Medical Journal of Australia*, **175**(4), 199–201. 16, 117

McLaughlin, Tara, Heat, Andrew, Bucholz, Kathleen, Madden, Pamela, Bierut, Lian, Slutske, Wendy, Dinwiddie, Stephen, Statham, Dixie, Dunne, Michael, & Martin, Nicholas. 2000. Childhood sexual abuse and pathogenic parenting in the childhood recollections of adult twin pairs. *Psychological Medicine*, **30**, 1293–1302. 22

Michalos, Alex C. 1991. *Global report on student well-being.* Vol. 1: Life-satisfaction and happiness. NY. Springer. 108

Michalos, Alex C, & Kahlke, P. Maurine. 2010. Stability and Sensitivity in Perceived Quality of Life Measures: Some Panel Results. *Social Indicators Research*, **98**(403). 106

Misheva, Violeta. 2014. Does victimization really increase offending? *working paper.*

Misheva, Violeta. 2015a. Do the 'young ones' in class need more support? *working paper.*

Misheva, Violeta. 2015b. What determines emotional well-being? The role of traumatic events: Evidence using twin data. *Journal of Happiness Studies, first online 6 October.*

Misheva, Violeta, Webbink, Dinand, & Martin, Nicholas. 2014. The Effect of Child Maltreatment on Illegal and Problematic Behavior. New Evidence on the 'cycle of violence' using twins data. *Working paper*.

Mouzos, Jenny, & Makkai, Toni. 2004. Women's experiences of male violence: Findings from the Australian component of the International Violence Against Women Survey (IVAWS). *Canberra, ACT: Australian Institute of Criminology, Research and Public Policy Series*, **56.** 16

Muehlenweg, Andrea M., & Puhani, Patrick A. 2010. The Evolution of the School-Entry Age Effect in a School Tracking System. *The Journal Of Human Resources*, **45**(2), 407–438. 3, 70

Muffels, Ruud, & Kemperman, Bauke. 2011. Does a Better Job Match Make Women Happier? Work Orientations, Work-Care Choices and Subjective Well-Being in Germany. *SOEP Paper*, **361**. 109

Mullen, Paul E, Martin, Judy L, Anderson, Jessie C, Romans, Sarah E, & Herbison, G Peter. 1993. Childhood Sexual Abuse and Mental Health in Adult Life. *British Journal of Psychiatry*, **163**, 721–732. 12

Myers, David G, & Diener, Ed. 1995. Who is happy? *Psychological Science*, **6**(1), 10–19. 106

Nelson, Elliot C, Heath, Andrew C, Lynskey, Michael T, Bucholtz, Kathleen K, Madden, Pamela A F, Statham, Dixie J, & Martin, Nicholas G. 2006. Childhood sexual abuse and risks for licit and illicit drug-related outcomes: a twin study. *Psychological Medicine*, **36**, 1473–1483. 6, 13

Nelson, Elliot C, Lynskey, Michael T, Heath, Andrew C, Madden, Pamela A F, & Martin, Nicholas G. 2010. A family study of adult twins with and without a history of childhood abuse: stability of retrospective reports of maltreatment and associated family measures. *Twin Research and Human Genetics*, **13**(2), 121–130. 37

Nelson, Henry B., & Martin, Catherine A. 1985. Increased child abuse in twins. *Child Abuse and Neglect*, **9**, 501–505. 16

Nettle, Danielle. 2005. *Happiness: The Science behind Your Smile*. Oxford University Press, Oxford, UK. 106

Neve, Jan-Emmanuel De, Christakis, Nicholas K, Fowler, James H, & Frey, Bruno. 2012. Genes, economics, and happiness. *Journal of Neuroscience, Psychology, and Economics*, **5**(4), 193–211. 104, 118

Okulicz-Kozaryn, Adam. 2010. Religiosity and Life Satisfaction Across Nations. *Mental Health, Religion and Culture*, **13**(155). 107

Okun, Morris, & Stock, William A. 1987. Correlates and components of subjective well-being among the elderly. *Journal of Applied Gerontology*, **6**, 95–112. 106

Oster, Emily. 2014. Unobservable Selection and Coefficient Stability: Theory and Evidence. *NBER working papers*, **19054.**

Ousey, Graham, Wilcox, Pamela, & Fisher, Bonnie S. 2011. Something old, something new: revisiting competing hypotheses of the victimization-offending relationship among adolescents. *Journal of Quantitative Criminology*, **27**, 53–84. 2, 42, 44, 45, 46, 47

Pajer, Kathleen A. 1998. What happens to 'bad' girls: A review of the adult outcomes of antisocial adolescent girls. *American Journal of Psychiatry*, **155**, 862–870. 11, 20

Pavot, William, & Diener, Ed. 1993. Review of the satisfaction with life scale. *Psychological Assessment*, **5**, 164–172. 114

Paxson, Christina, & Waldfogel, Jane. 1999. Parental resources and child abuse and neglect. *American Economic Review*, **89**(2), 239–44. 24

Paxson, Christina, & Waldfogel, Jane. 2002. Work, Welfare, and Child Maltreatment. *Journal of Labor Economics*, **20**(3), 435–74. 24

Peiro, Amado. 2006. Happiness, Satisfaction and Socio- Economic Conditions: Some International Evidence. *The Journal of Socio-Economics*, **35**(348). 105, 106

Putnam, Frank W. 2003. Ten-Year Research Update Review: Child Sexual Abuse. *Journal of the American Academy of Child and Adolescent Psychiatry*, **42**(3), 269–278. 12

Robarge, Joyce P., Reynolds, Zelta B., & Groothuis, Jessie R. 1982. Increased child abuse in families with twins. *Research in Nursing and Health*, **5**(4), 199–203. 16

Rosenbaum, Paul, & Rubin, Donald B. 1983. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society*, **45**(2), 212–218. 56

Ryan, Richard M, Huta, Veronika, & Deci, Edward L. 2008. Living well: A self-determination theory perspective of eudaimonia. *Journal of Happiness Studies*, **9**, 139–170.

Sampson, Robert J, & Lauritsen, Janet L. 1990. Deviant Lifestyles, Proximity to Crime, and the Offender-Victim Link in Personal Violence. *Journal of Research in Crime and Delinquency*, **27**(2.). 2, 43

Schreck, Christopher. 1999. Criminal victimization and low self-control: an extension and test of a general theory of crime. *Justice Quarterly*, **16**(3), 633–654. 46

Schreck, Christopher, Stewart, Eric A., & Fisher, Bonnie S. 2006. Self-control, Victimization and their Influence on Risky Lifestyles: A Longitudinal Analysis Using Panel Data. *Journal of Quantitative Criminology*, **22**, 319–340. 42

Schreck, Christopher, Stewart, Eric A., & Osgood, D Wayne. 2008. A Reappraisal of the overlap of violent offenders and victims. *Criminology*, **46**(4.). 44

Schwarz, Norbert, & Clore, Gerald L. 1983. Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology*, **45**(3), 513–523. 114

Searight, Russell, Rottnek, Fred, & Abby, S L. 2001. Conduct disorder: diagnoses and treatment in primary care. *American Family Physician*, **63**(8), 1579–1588. 20

Seepersad, Randy. 2010. Review of the roots of youth violence: Literature review. *The Review of the Roots of Violence*, **5**. 46

SEO. 2007. De costen van criminaliteit [The costs of crime]. *Report No*, **971.** 43

Sham, Pak C, O'Callaghan, Evelyn, Takei, Naoko, Murray, Gordon, Har, Ed, & Murray, Raymond. 1994. Schizophrenia following pre-natal exposure to influenza epidemics between 1939 and 1960. *The British Journal of Psychiatry*, **160**, 461–466. 76

Shields, Michael A ., Price, Stephen Wheatley, & Wooden, Mark. 2009. Life satisfaction and the economic and social characteristics of neighborhoods. *Journal of Population Economics*, **22**, 421–443. 105, 107

Short, James F., & Nye, F. Ivan. 1957. Reported Behaviour as a Criterion of Deviant Behaviour. *Social Problems*, **5**, 207–13. 20

Simons, Ronald L., Wu, Chyi-In, Johnson, Christine, & Conger, Rand D. 1995. A test of various perspectives on the intergenerational transmission of domestic violence. *Criminology*, **33**(1), 141–172. 11, 20

Singer, Simon I. 1981. Homogenous victim-offender populations: A review and some research implications. *Journal of Criminal Law and Criminology*, **71**, 779–788.

Singer, Simon I. 1986. Victims of serious violence and their criminal behavior: Subcultural theory and beyond. *Violence and Victims*, **1**, 61–70.

Smith, David, & Ecob, Russel. 2007. An Investigation into causal links between victimization and offending in adolescents. *The British Journal of Sociology*, **58**. 2, 42, 43

Sobel, Russel S, & Osaba, Brian J. 2009. Youth gangs as Pseudo-Governments: Implications for Violent Crime. *Southern Economic Journal*, **74**(4), 996–1018. 47

Sparks, R. 1982. *Research on Victims of Crime*. US Governmental Printing Office, Washington, DC.

Stipek, Deborah. 2002. At what age should children enter kindergarten? A question for policy makers and parents. *Social Policy Report*, **16**(1), 3–16. 3, 70

Stock, William A, Okun, Morris A, & Benin, Mary. 1986. Structure of subjective well-being among the elderly. *Psychology and Aging*.

Stutzer, Alois, & Frey, Bruno. 2006. Does marriage make people happy, or do happy people get married? *The Journal of Socio-Economics*, **35**, 326–347. 106, 120

Tella, Rafael Di, MacCulloch, Robert J., & Oswald, Andrew J. 2001. Preferences over inflation and unemployment: evidence form surveys of happiness. *The American Economic Review*, **91**(1). 121

Tellegen, Alike, Bouchard, David T. Lykken Thomas J, Wilcox, Kimerly J, Segal, Nancy L, & Rich, Stephen. 1988. Personality Similarity in Twins Reared Apart and Together. *Journal of Personality and Social Psychology*, **54**(6), 1031–1039. 103

Thoits, Peggy, & Hannan, Michael. 1979. Income and psychological distress: The impact of an income-maintenance experiment. *Journal of Health and Social Behavior*, **20**, 120–138. 108

Veenhoven, Ruut. 1983. The Growing Impact of Marriage. *Social Indicators Research*, **12**(49).

Veenhoven, Ruut. 1984. *Conditions of happiness*. D. Reidel, Dordrecht.

Veenhoven, Ruut. 2008. Healthy Happiness:Effects of happiness on physical health and consequences for preventive health care. *Journal of Happiness Studies*, **9**, 449–469. 107

Veenhoven, Ruut. 2009. *How do we assess how happy we are?* A. K. and Radcliff, B. (eds), Happiness, Economics and Politics: Towards a multi-disciplinary approach, Edward Elger Publishers, Cheltenham UK. 103

Ventegodt, Soren. 1999. A Prospective Study on Quality of Live and Traumatic Events in Early Life - A 30 Year Follow-Up. *Child: Care, Health and Development*, **25**(213). 109

Ward, Shanon, & Williams, Jenny. 2014. Does Juvenile Delinquency Reduce Educational Attainment? *working paper series (22 July 2014)*. 55

Webbink, Dinand, Vujic, Suncica, Koning, Pierre, & Martin, Nicholas. 2012. The Effect of Childhood Conduct Disorder on Human Capital. *Health Economics*, **8**, 928–945. 19

Webbink, Dinand H, Koning, Pierre, Vujic, Suncica, & Martin, Nicholas. 2013. Why are criminals less educated than non-criminals? *Journal of Law, Economics, and Organization*, **29**(1), 115–144. 20

Widom, Cathy S. 1989. Does Violence Beget Violence? A critical examination of the literature. *Psychological Bulletin*, **106**(1), 3–28. 2, 7, 11, 12

Widom, Cathy S. 1992. The Cycle of Violence. *Research in Brief*. 2, 7

Widom, Cathy S, & Maxfield, Michael G. 2001. An Update on the Cycle of Violence. *Research in Brief*. 2, 7

Williams, Linda M. 1994. Recall of childhood trauma: A prospective study of women's memories of child sexual abuse. *Journal of Consulting and Clinicl Psychology*, **62**, 1167–1176. 37

Williams, Rober L., Eyring, Margot, Gaynor, Pat, & Long, James D. 1991. Development of views of life scale. *Psychology and Health*, **5**, 165–181. 114

Wilson, Warner R. 1967. Correlates of avowed happiness. *Psychological Bulletin*, **67**, 294–306. 115

Winkelmann, Liliana, & Winkelmann, Rainer. 1998. Why are the unemployed so unhappy? Evidence from panel data. *Economica*, **65**(257), 1–15. 112

Wittebrood, Karin, & Nieuwbeerta, Paul. 1999. Wages of sin? The link between offending, lifestyle and violent victimization. *European Journal on Criminal Policy and Research*, **7**, 63–80. 50

Wittebrood, Karin, & Nieuwbeerta, Paul. 2000. Criminal Victimization During One's Life Course: The Effects of Previous Victimization and Patterns of Routine Activities. *Journal of Research in Crime and Delinquency*, **37**(91.).

Wittebrood, Karin, & ter Voert, M.J. 2007. *Netherlands Survey on Crime and Law Enforcement: Documentation of the Survey.* NSCR, Leiden. 50

Wolfgang, Marvin, & Ferracuti, Franco. 1967. *The subculture of violence: towards and integrated theory in criminology.* London: Tavistock. 2, 42, 45

Wooldridge, Jeffrey M. 2012. *Econometric analysis of cross section and panel data.* the MIT Press, Cambridge, Massachussets.

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus University Rotterdam, University of Amsterdam and VU University Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

597 T.A. MAKAREWICZ, *Learning to Forecast: Genetic Algorithms and Experiments*

598 P. ROBALO, *Understanding Political Behavior: Essays in Experimental Political Economy*

599 R. ZOUTENBIER, *Work Motivation and Incentives in the Public Sector*

600 M.B.W. KOBUS, *Economic Studies on Public Facility use*

601 R.J.D. POTTER VAN LOON, *Modeling non-standard financial decision making*

602 G. MESTERS, *Essays on Nonlinear Panel Time Series Models*

603 S. GUBINS, *Information Technologies and Travel*

604 D. KOPÁNYI, *Bounded Rationality and Learning in Market Competition*

605 N. MARTYNOVA, *Incentives and Regulation in Banking*

606 D. KARSTANJE, *Unraveling Dimensions: Commodity Futures Curves and Equity Liquidity*

607 T.C.A.P. GOSENS, *The Value of Recreational Areas in Urban Regions*

608 Ł.M. MARĆ, *The Impact of Aid on Total Government Expenditures*

609 C. LI, *Hitchhiking on the Road of Decision Making under Uncertainty*

610 L. ROSENDAHL HUBER, *Entrepreneurship, Teams and Sustainability: a Series of Field Experiments*

611 X. YANG, *Essays on High Frequency Financial Econometrics*

612 A.H. VAN DER WEIJDE, *The Industrial Organization of Transport Markets: Modeling pricing, Investment and Regulation in Rail and Road Networks*

613 H.E. SILVA MONTALVA, *Airport Pricing Policies: Airline Conduct, Price Discrimination, Dynamic Congestion and Network Effects*

614 C. DIETZ, *Hierarchies, Communication and Restricted Cooperation in Cooperative Games*

615 M.A. ZOICAN, *Financial System Architecture and Intermediation Quality*

616 G. ZHU, *Three Essays in Empirical Corporate Finance*

617 M. PLEUS, *Implementations of Tests on the Exogeneity of Selected Variables and their Performance in Practice*

618 B. VAN LEEUWEN, *Cooperation, Networks and Emotions: Three Essays in Behavioral Economics*

619 A.G. KOPÁNYI-PEUKER, *Endogeneity Matters: Essays on Cooperation and Coordination*

620 X. WANG, *Time Varying Risk Premium and Limited Participation in Financial Markets*

621 L.A. GORNICKA, *Regulating Financial Markets: Costs and Trade-offs*

622 A. KAMM, *Political Actors playing games: Theory and Experiments*

623 S. VAN DEN HAUWE, *Topics in Applied Macroeconometrics*

624 F.U. BRÄUNING, *Interbank Lending Relationships, Financial Crises and Monetary Policy*

625 J.J. DE VRIES, *Estimation of Alonso's Theory of Movements for Commuting*

626 M. POPŁAWSKA, *Essays on Insurance and Health Economics*

627 X. CAI, *Essays in Labor and Product Market Search*

628 L. ZHAO, *Making Real Options Credible: Incomplete Markets, Dynamics, and Model Ambiguity*

629 K. BEL, *Multivariate Extensions to Discrete Choice Modeling*

630 Y. ZENG, *Topics in Trans-boundary River sharing Problems and Economic Theory*

631 M.G. WEBER, *Behavioral Economics and the Public Sector*

632 E. CZIBOR, *Heterogeneity in Response to Incentives: Evidence from Field Data*

633 A. JUODIS, *Essays in Panel Data Modelling*

634 F. ZHOU, *Essays on Mismeasurement and Misallocation on Transition Economies*

635 P. MULLER, *Labor Market Policies and Job Search*

636 N. KETEL, *Empirical Studies in Labor and Education Economics*

637 T.E. YENILMEZ, *Three Essays in International Trade and Development*

638 L.P. DE BRUIJN, *Essays on Forecasting and Latent Values*

639 S. VRIEND, *Profiling, Auditing and Public Policy: Applications in Labor and Health Economics*

640 M.L. ERGUN, *Fat Tails in Financial Markets*

641 T. HOMAR, *Intervention in Systemic Banking Crises*

642 R. LIT, *Time Varying Parameter Models for Discrete Valued Time Series*

643 R.H. KLEIJN, *Essays on Bayesian Model Averaging using Economic Time Series*

644 S. MUNS, *Essays on Systemic Risk*

645 B.M. SADABA, *Essays on the Empirics of International Financial Markets*

646 H. KOC, *Essays on Preventive Care and Health Behaviors*