Essays in microeconomics with applications in education, health and crime

Franz Georg Westermaier

Dissertation submitted to the Hertie School of Governance

in partial fulfilment of the requirements for the degree of

Doctor rerum politicarum (Dr. rer. pol.)

in the Doctoral Programme in Governance

Berlin, 2016

First reviewer

Prof. Dr. Christian Traxler

Herite School of Governcance

& Max Planck Institute for Research on Collective Goods (MPI Bonn)

Second reviewer

Prof. Dr. Peter Haan

Freie Universität Berlin

& German Institute for Economic Research (DIW Berlin)

Acknowledgements

First of all, I would like to gratefully acknowledge the supervision of Professor Christian Traxler, who provided me with a perfect mix of instructive criticism and constant encouragement. This thesis has gained substantially from his invaluable comments and suggestions.

During the past four years, I have also received a lot of support from my colleagues at the Hertie School of Governance within internal seminars. Many thanks are also due to Felix Albrecht, my colleague at the Philipps-University of Marburg, for fruitful and stimulating discussions even beyond the scope of this thesis.

While preparing this thesis, I have benefited enormously from collaborations with Andrea Mühlenweg, Brant Morefield and Ansgar Wohlschlegel as well as my supervisor Christian Traxler. It was – and continues to be – a pleasure to work with you!

Financial support from the Hertie Foundation is gratefully acknowledged.

Finally, I want to thank my parents Gerda and Fritz Westermaier, my siblings Carola and Max Westermaier as well as my flatmates and friends Philipp Gübler, Jane Lechler, Gianina Meneses, Marcos Moreno, Mark Praznik, Irina Rogozhina, Isabel Urrutia and Margarita for their support and tolerance during all these years.

Franz Georg Westermaier

Berlin, January 2016

Contents

List of Tables

List of Figures

Preface

I herewith present my dissertation in the field of microeconomics, which demonstrates my ability to carry out advanced independent scientific work and provides an advancement of scientific knowledge in this and related fields. Simple changes in an institutional framework can have huge impacts for the human society which makes it important to analyze them and define their consequences to create a surplus for the society.

In the following dissertation, I present applications how policy regulations and external events generate behavioral responses with a focus on indirect effects, e.g. from health on education and education on crime. All chapters, while not exclusively, follow an empirical approach and show effects in important fields of human society. Beyond the research methods from microeconomics, this dissertation tackles the field of health, early childhood development, secondary education, adolescent drug-abuse and law enforcement. Several identification strategies are used to cope with endogeneity and deliver causal effects between the fields of education, health and crime.

Every chapter of this dissertation evaluates institutional changes on behavioral responses in different stages of life: early-childhood, adolescence and adulthood. Chapter 1 focuses on early-childhood and relies on exogenous changes of the family's constraints affecting the most important institution for a young child, her family. Chapter 2 analyzes an educational reform effect on adolescents' behavior in substance abuse and crime. Chapter 3 makes use of the design of an institutional framework itself to analyze how human behavior is affected by it.

Chapter 1 examines the importance of parental health in the development of child behavior during early childhood.[1] With respect to the formation of children's cognitive and non-cognitive skills, the early childhood is a crucial stage of skill development since it is the foundation for all other skills that are achieved later on in life. Furthermore, the family or specifically changes in family's budget constraints are important because they account for most of the environment which is influencing children in their early childhood.

This analysis is based on child psychometric measures from a longitudinal German dataset, which tracks mothers and their newborns up to age six. We identify major changes in parental health (shocks) and control for a variety of initial characteristics of

[1] This chapter is co-authored with Andrea M. Mühlenweg and Brant Morefield.

the child including prenatal conditions. The results are robust to placebo regressions of health shocks that occur after the outcomes are measured. Furthermore, we can rule out that the measured effects of parental health shocks are driven by other serious events like divorce, job-losses etc.. While these events might exert additional effects on children's non-cognitive skills they are not biasing our main results. Our findings point to negative effects of maternal health shocks on children's emotional symptoms, conduct problems and hyperactivity. We estimate that maternal health shocks worsen outcomes by as much as 0.9 standard deviations. In contrast, paternal health seems to be less relevant to children's behavioral skills.

Chapter 2 focuses on how student's behavior in crime and illegal substance abuse can be affected by an educational reform. Adolescent students are in their crucial stage to gain drug experiences or engage in criminal activity. In contrast to the previous literature which focuses on high-risk groups with respect to crime, the analyzed reform affected high performing students in education with higher opportunity costs to engage in crime.

During the last decade, a major educational reform in Germany reduced the academic high school duration by one year while keeping constant the total number of instructional hours before graduation. The instructional hours from the eliminated school year shifted to lower grade levels, which increased the time younger students spend at school. This study explores the impact of the reform on youth crime rates and substance abuse using administrative police crime statistics, administrative student enrollment data, and a student drug survey.

The staggered implementation of the reform in different *Länder*-age-groups allows for a difference-in-difference approach. I find that the reform resulted in a decline in crime rates, which is almost exclusively driven by a reduction in violent crime and illegal substance abuse. Regarding the latter, the rate of illegal cannabis consumption strongly declined; however, no significant effect is detected on cannabis dealers or the consumption of other illegal drugs.

Declining cannabis addiction of reform-affected students is also documented in a repeated student drug survey. The survey evidence further suggests that decreased cannabis consumption was not driven by a shift of consumption into 'school hours'. The results point to an 'incapacitation' effect of schooling due to the intensified curriculum at lower grade levels: The extent to which the effect is driven by incapacitation in school classrooms or 'self-incapacitation' at home to cope with the higher study workload remains unclear.

Chapter 3 analyzes the reaction of adult behavior within a penalty scheme for speeding.[2] In comparison to the first two chapters which are focusing on effects in the long- or medium- run, the last chapter studies responses to an institutional setting where the individual decision making process is performed in a really short time horizon.

The paper studies drivers' responses to a 'notched' penalty scheme in which speeding penalties are stepwise increasing with discontinuous jumps at several speed levels.

We first present survey evidence suggesting that drivers in Germany are very well aware of the notched penalty structure. Based on a simple analytical framework we then analyze the impact of the notches on drivers' optimal speed choices. The model's predictions are then confronted with data on more than 150,000 speeding tickets from violations occurring on the German Autobahn.

The data provide evidence on bunching: many drivers who speed still stay below a penalty notch. For major speed limit violations, however, we do not detect bunching. The paper further explores the impact of a recent policy reform and discusses the normative implications of our findings. The results from our positive analysis also carry implications for the normative debate on optimal speed limit enforcement and optimal penalties in general.

All chapters of this dissertation show impacts of different institutional settings which shape human's behavior in different stages of their lifetime. While the first two chapters focus on major changes in institutions of young children and adolescents, the third chapter shows that even small details of an institutional design can have a huge impact on the design-making process of (boundedly) rational adults.

[2] This chapter is co-authored with Christian Traxler and Ansgar Wohlschlegel.

Chapter 1

Parental Health and Child Behavior: Evidence from parental health shocks

Co-authored with Andrea M. Mühlenweg & Brant Morefield

1.1 Introduction

This study examines the importance of parental health in the development of children's non-cognitive skills during early childhood. We draw on standard psychometric measures in young children, covering emotional symptoms, conduct problems, hyperactivity, and peer-relationship problems. For simplicity, we refer to these items as *behavioral skills*. To our knowledge, there is limited evidence regarding parental health effects on the development of child behavioral skills to date. At the same time, these respective skills are known to be important components of human capital, yielding improved education and labor market outcomes (e.g. Cunha and Heckman 2008).

From a theoretical point of view, one would expect that poor parental health affects a family's budget constraints altering parental investment in children's skills. For instance, a less healthy adult may face tighter budget constraints because she is forced to spend down family wealth (Wu 2003), is less productive in performing chores or in the labor market (Podor and Halliday 2012) and earns a lower income (Currie and Madrian 1999).[1] Poor parental health may also reduce both the productivity and the amount of time parents spend with their children (cf. Ruhm 2004 and Morefield 2010 for empirical evidence). Additionally, poor parental health may have a direct impact as it can traumatize or even stigmatize children (Coneus et al. 2012). In any case, the nature of the technology of skill formation suggests potentially long-lasting effects from even temporary reductions in investment in skill development (Cunha et al. 2006). For instance, skill formation includes sensitive periods, wherein investments are more productive, and remediation of reduced investment during these periods is more costly in later periods. As such, we may expect differing investment periods to exert different impacts; our study distinguishes between parental health in two periods of early childhood, these being from ages 0 to 3 and from ages 3 to 6.

The identification of parental health effects on child outcomes is challenged by the fact that parental health may be endogenous to the formation of children's skills. For example, genetic dispositions for specific (mental) health conditions may affect both parental health and children's psychopathological outcomes. In addition, small, but permanent, depreciations in the parental health status evolving over time could be a result of individual decisions such as deleterious living conditions, which may also affect investments in children's skills. Therefore, we use sudden changes (shocks) in parental health instead of current health status as an identifying source of variation. We assume that a sudden drop in health is less likely to be determined by the individual's decision making process and, therefore, more exogenous. Similarly, we expect genetically pre-determined health conditions to exert more permanent health effects over time. In contrast, potential reasons for major period-to-period shocks in observed health measures are accidents or the (unexpected) onset of a physical handicap of disease (cf. Riphan 1999). However, we cannot rule out that there are cases where a sudden onset of a severe health condition will be unexpected and endogenous in the sense of our identification strategy. Therefore, we additionally control for a variety of mother and child variables, including prenatal (health) conditions and also include current household events such as family disruption or job loss in a robustness check. Our shock-based approach corresponds to a common strategy used to identify effects of health on socio-economic measures (e.g. Smith 1999; Smith 2005; Riphan 1999; Jones et al. 2010). To test the identifying assumption of our model, we present several

[1] Adda et al. (2009), Riphahn (1999) and Smith (2004) provide further evidence on income changes due to health shocks.

robustness checks including results from placebo regressions where we regress child outcomes on *future* parental health shocks.[2]

This paper follows recent empirical studies that demonstrate the importance of early life events on human capital development. While the general importance of home investments during the early part of a child's life has been shown (e.g. Todd and Wolpin 2007; Blomeyer et al. 2009 based on German data), attempts to quantify the effects of commonly experienced household shocks are limited. A related strand of literature examines the effects of changes in family structure on children's outcomes (cf. Ribar 2004 for a review with a focus on marriage). Additional studies on changes in family background and child outcomes consider parental employment and life satisfaction. For Germany, Berger et al. (2010) show that permanent maternal unemployment negatively affects children's adaptive behavior. Berger and Spieß (2011) provide evidence on the positive impacts of maternal life satisfaction on children's verbal and socio-emotional skills.[3]

Besides this literature, there are few papers in the economics literature that seek to identify the causal effects of parental health on child outcomes with a focus mainly on child educational outcomes:[4]

Johnson and Reynolds (2013) analyze the effects of household members' hospitalization on the educational attainment of adolescents. Controlling for pre-hospitalization background, they find negative effects of hospitalizations lasting one week or longer. Other household members' hospitalizations impact the probability of children completing high school, attending college and receiving a university degree. Negative effects are

[2]Note that our research question differs from the literature on the in-utero influences of maternal health on child outcomes (cf. Currie and Almond 2011 for a review). We are interested in how child behavior is affected by variations in parental health *in the early years of childhood* (after birth). Therefore, initial maternal and child health (at birth) are used as control variables in our empirical approach.

[3]See Currie and Almond (2011) for an international review of further studies. In addition, studies on parental death might be considered as the most extreme health shock (cf. Adda et al. 2011 for a recent paper and a review of the evidence). They find small negative effects on skill development and somewhat lower earnings later in life for the affected children. Based on a difference-in-differences approach, Senne (2014) provides strong evidence that adult mortality has short- and long-run negative impacts on children's educational outcomes.

[4]Our work also relates to studies that examine the impact of maternal psychiatric illness, the most common of which are depression or substance abuse and smoking during pregnancy, on children's outcomes. The results consistently show that children of depressed mothers fare worse on a wide range of outcomes, including the development of cognitive and motor skills (Petterson and Albers 2001), problem behavior (Frank and Meara 2009), and social behavior (Kim-Cohen et al. 2005). Farahati et al. (2003) find that parental psychiatric illness is associated with a lower probability of high school graduation among children. Balsa (2008) provides evidence that parental problem drinking exerts negative impacts on children's labor market performance later in life.

more pronounced for male adolescents, and there are differential effects by the birth order and gender of older siblings. Andrews and Logan (2010) examine whether parental health status accounts for test-score gaps between school-age children of different ethnicities. The authors find that controlling for a large set of parental health measures reduces the test-score gap between Blacks and Whites (Hispanic and Whites) by 17 percent (10 percent). Sun and Yao (2010) draw on a panel of rural Chinese households to analyze how parental health shocks impact school-aged children's educational attainment. According to their findings, parental health shocks negatively affect younger children's outcomes, while there is no measurable impact on students in secondary school. The study suggests that the educational effects are mainly due to limited financial resources.

While this previous literature focuses on educational outcomes and cognitive skills, there is little evidence on parental health effects on behavioral outcomes. Morefield (2010) is most closely related to our work. Besides cognitive (mathematical) skills, the study examines a measure of children's problem behavior which is similar to the aggregated psychometric measure we use in our analysis. The paper analyzes whether the onset of parental health conditions relates to an increase in children's problem behavior. In contrast to our study, Morefield (2010) does not consider parental health shocks but the reported onset of specific health conditions limiting parents' usual daily activities. The study finds that these health conditions at ages 5-9 are related to significant increases in children's problem behaviors but not at younger or older ages (up to age 18). And, there are no significant effects on the children's cognitive skill measure.

Beyond the evidence for an aggregated psychometric measure, our study provides detailed results for a battery of behavioral skills. Our findings show significant negative effects of maternal health shocks on the aggregated behavioral score and on children's emotional symptoms, conduct problems, and hyperactivity. We do not find such effects in "placebo regressions" where we estimate the impact of parental health shocks that occur after a score for non-cognitive skills in early childhood is measured. The results are also robust to controlling for additional household events that may drive a spurious relationship (i.e. job loss, family disruption, and income loss). While an examination of the mechanisms that link parental health and child outcomes is beyond the scope of this paper, an additional analysis that includes family income suggests that financial investments in children's skills development are not the main drivers of the observed effects. This is consistent with previous research on household income on children's outcomes, which implies that quality or quantity of parental time are more important than financial resources (e.g. Propper et al. 2007; Violato et al. 2011; Dooley and Stewart 2007).

To our knowledge, the current study is the first that examines parental health shocks and children's skills outcomes based on German data. Germany differs from the US in many institutional aspects that may be relevant to the translation of parental health problems to child behavior effects. For example, the availability and level of paid sick leave policies are much more generous in Germany and assure that employees receive financial coverage for longer periods requiring time off from work (cf. Heymann et al. 2009 for an international comparison).[5] Therefore, one might assume that tighter budget constraints due to parental illness will be a less severe problem in Germany compared to the US. Similarly, parental leave and parental benefit regulations are relatively generous in Germany. Parental leave predominantly taken by mothers is generally guaranteed for three years, prolonging to six years if a second child is born. In line with a predominant male breadwinner model, most of the employed German mothers work part time to have more time for child-rearing activities (cf. Ciccia and Verloo 2012). As a stylized fact, according to OECD figures, the share of families with two full-time working parents amounts to less than 20 percent in German two-parent families with children under age 14; this proportion is roughly 70 percent for comparable families in the US (OECD 2012).

These stylized differences with German mothers predominantly taking the role of children's caregivers and the breadwinner's health being less important with respect to sustaining family income may point to a potentially higher importance of maternal health in Germany as compared to the US. Our empirical analyses distinguish between maternal and paternal health shocks and find that maternal health shocks are especially harmful. This result differs from evidence in the above mentioned study by Morefield (2010) based on US data. Morefield (2010) fails to identify statistically significant impacts of maternal health conditions on child outcomes while there is evidence that paternal health conditions negatively impact child behavior. However, the author notes the low precision of estimates conditional on parent's gender due to a small sample

[5] Heymann et al. (2009) systematically assess paid-sick-pay and paid-sick leave policies in 22 countries. While the US ranks worst, Germany is ranked as one of the top five countries (based on coverage for full-time equivalent working-days for median earners). In general, sick-pay in Germany amounts to 100 percent of an employee's previous salary for up to six weeks. After this period, the public health insurance companies cover about 70 percent of the individual's regular gross income for up to 78 weeks within each three year period.

size (cf. the Conclusion section of Morefield 2010).[6] For the aggregated results not conditioned on parent's gender our findings are in line with Morefield (2010): In both studies, parental health events are especially harmful in the observed age groups that include school entry age (here: age 3-6, Morefield: age 5-9).[7]

The remainder of our paper proceeds as follows. Section 1.2 introduces the data used and provides descriptive evidence. Section 1.3 discusses our empirical approach. Section 1.4 presents and discusses our findings together with evidence from several robustness checks. Section 1.5 concludes.

1.2 Data and Descriptive Evidence

Our empirical analysis is based on the German Socio-Economic Panel Study (SOEP).[8] The SOEP is a representative panel study that records annual information on approximately 20,000 adults that live in approximately 12,000 households. While several indicators of health are collected over time in the SOEP, the most consistently fielded health-related question gathers self-reported health satisfaction − coded from zero (completely dissatisfied) to ten (completely satisfied). In addition, as a more objective measure, we use yearly information on the number of nights spent in hospital.[9]

In 2003, the SOEP began collecting the so-called "mother-and-child data" which we use for our study. The database contains information on newborn children (i.e. younger than 1.5 years) and their mothers. Further information on the children is collected when

[6]Because of different samples and estimation strategies, our findings are not directly comparable to Morefield (2010). Particularly, Morefield (2010) considers a sample of all children up to age 18 when comparing paternal and maternal health events. This analysis does not differentiate between age groups while our results relate to young children up to age three and from age three to age six respectively. Morefield (2010) examines changes in age-adjusted measures conditional on the reported onsets of specific diagnoses and parental health limitations instead of health shocks while controlling for parental and child background variables including children's home learning environment.

[7]Ermisch et al. (2012) use the same data and aggregated behavioral measure as in our paper. They find that multiple maternal relationship changes (due to separations or new partners) are related to an increase in the aggregated measure by about 0.4 standard deviations for children observed when they are about six years old. In terms of magnitude this also corresponds to the size of the estimated parental health effects on the aggregated outcome for children aged 3-6 in our study.

[8]Cf. Wagner et al. (2007) for an overview and introduction to the dataset.

[9]Further SOEP health questions include health limitations, handicaps, chronic diseases, health deteriorations, days of sick leave, current state of health, number of hospital visits, medical care after work accidents, and any doctor's visits in the previous three months. However, the respective information is only collected biannually, conditional on employment and/or covers limited periods of time. In sum, these measures are not appropriate for our estimation strategy, which relies on year-to-year changes in health measures. In addition, the sample size does not allow for differentiating the specific health conditions that underlie the observed health shocks.

they are about three years old (23 years) and again when they are about six years old (56 years). One feature of the mother-and-child data is the measurement of children's non-cognitive outcomes. We use a modified version of the Strength and Difficulties Questionnaire (SDQ) measured when the children are approximately six years old. The SDQ is a standard psychometric measure and is based on mothers' assessments of their children's behavior and socio-emotional skills (Goodman 1997) and is commonly used in psychopathological screening (e.g. Becker et al. 2006; Achenbach et al. 2008). We show results for two aggregated measures derived from the SDQ: *children's socio-emotional behavior (SEB)* and the *pro-social behavior score (PBS).*

The SEB indexes four dimensions of non-cognitive skills, for which we provide detailed evidence, including *emotional symptoms*, *conduct problems*, *hyperactivity and inattention*, and *peer-relationship problems.*

The PBS is based on maternal judgment of children's thoughtfulness, cooperation, and helpfulness.[10] For ease of interpretation, we have altered the measures so that lower scores consistently represent worsened behavioral outcomes and are standardized (z-scores).

In addition to the SDQ, one additional skill measure is used for our robustness checks: the Vineland Adaptive Behavior Scale (VAB) which we observe for the sample of three-year-olds. The VAB is an index of children's non-cognitive skills based on parental reports of children's verbal skills, activities of daily living, motor skills, and social skills.[11] We use the VAB to estimate a "placebo regression" wherein we estimate the effects of parental health shocks that occur both prior to and after the VAB is collected. Under the identifying assumption, health shocks that occur after the VAB is collected should not be correlated with child scores in our model.

The SOEP includes data on 703 children from birth to age six. We exclude individuals with missing information on the mother's age when she delivered (23 observations), the week of gestation at the time of the child's birth (15), and child's birth weight (2). Of the remaining 663 children, we observe the SDQ (VAB) outcome for 639 (634) children.

[10] Table 1.5 in the Appendix provides the detailed questions underlying all the SDQ items.

[11] Table 1.6 in the Appendix lists all questions to gather the VAB score. However, we do not present more detail on the Vineland Scale since it is solely used to conduct a robustness check. The interested reader is referred to Schmiade et al. (2008) who summarize the use of the Vineland Scale as a measure of child development in the SOEP.

Table 1.1 provides the mean values of the children's skill measures (column 1) and means conditional on the parents' self-rated health satisfaction. Approximately 20 percent of parents are in "bad health" when their child is six years old, where "bad health" is defined as values from 0 to 5 on the eleven-point scale and "good health" values range from 6 to 10. Table 1.1 implies that children of healthier mothers have more favorable socio-emotional outcomes. The mean difficulty score is about 0.4 standard deviations higher for children whose mothers are in good health. Looking at the SDQ sub-scales suggests that having a less healthy mother is significantly related to a poorer emotional symptom score, a higher incidence of hyperactivity and conduct problems, and less favorable pro-social behavior. The means do not indicate significant differences in child behavior in relation to paternal health.[12] All SDQ scores imply less favorable outcomes for households without a father, but the difference is not statistically significant. We treat the mother's partner living in the household as a father. According to this definition, 21 percent of the sampled children are in single mother households.

[12]Table 1.1 also includes the mean Vineland scores that inform on children's adaptive behavior at age three. None of the observed differences are statistically significant conditional on parental health at age three (not shown) or age six (included in Table 1.1).

Table 1.1: Parental health satisfaction and child behavior

Variable	Overall Sample	Mother: "good health"	Mother: "bad health"	Father: "good health"	Father: "bad health"	Father in household	No Father in household
Overall difficulty score (R)	0.000	0.108	-0.310***	0.025	0.017	0.024	-0.087
(Age 6)	(0.045)	(0.043)	(0.090)	(0.050)	(0.087)	(0.044)	(0.091)
Emotional Symptoms (R)	0.000	0.076	-0.198***	0.031	-0.021	0.021	-0.076
(Age 6)	(0.045)	(0.043)	(0.090)	(0.050)	(0.093)	(0.044)	(0.090)
Hyperactivity (R)	0.000	0.111	-0.339***	0.008	0.058	0.018	-0.064
(Age 6)	(0.045)	(0.044)	(0.083)	(0.050)	(0.093)	(0.044)	(0.088)
Conduct problems (R)	0.000	0.071	-0.206***	0.012	0.002	0.010	-0.035
(Age 6)	(0.045)	(0.043)	(0.035)	(0.049)	(0.095)	(0.044)	(0.091)
Peer relationship problems (R)	0.000	0.024	-0.071	0.017	0.012	0.017	-0.063
(Age 6)	(0.045)	(0.047)	(0.086)	(0.050)	(0.095)	(0.045)	(0.086)
Pro-social behavior	0.000	0.048	-0.116**	0.024	0.012	0.025	-0.092
(Age 6)	(0.045)	(0.045)	(0.082)	(0.050)	(0.095)	(0.044)	(0.088)
Adaptive behavior	0.000	0.004	-0.021	-0.002	0.031	0.008	-0.045
(Vineland, age 3)	(0.040)	(0.043)	(0.102)	(0.049)	(0.095)	(0.044)	(0.085)
Observations (Age 6, SDQ and SEB Score)	639	483	151	401	101	502	137
Observations (Age 3, Vineland)	634	525	109	444	91	535	99

Notes: Means (and standard errors) of the respective variables. (R = reverse scale): Higher outcome scores always imply more favorable behavioral skills. * Marks the statistically significant difference between a parent in good health and a parent in bad health at the ten percent level, ** at the five percent and *** at the one percent level. We consider parents to be of "good" health if they rate their own health as good (from 6 to 10) on the eleven-point scale and to be of "bad" health if they report intermediate or bad health (values from 0 to 5). We do not observe health satisfaction for five mothers at the time of their childrens skill assessment. However, information on these five mothers is still used in the regression analysis because we observe their health history after they gave birth and also required the relevant child outcomes.
Source: Mother-and-child data of the German Socio-Economic Panel Study (SOEP), version 28. Own calculations.

1.3 Empirical Approach

The descriptive information in Table 1.1 shows less favorable non-cognitive outcomes for children whose parents suffer from bad health. However, this may be due to unobserved characteristics driving parental as well as child outcomes. Parents' and children's human capital are likely to be interrelated genetically and environmentally. For example a genetic disposition for a specific mental-health limitation might go along with a less favorable development of cognitive or behavioral skills. Therefore, in order to estimate the effects of parental health in the early years of a child's life, we examine parental health *shocks* instead of health levels. We define health shocks as sizeable year-to-year changes in an individual's self-reported health satisfaction or in the number of nights she spent in a hospital during the year. More specifically, we define a shock in health satisfaction as a year-to-year *drop* in health satisfaction of two or more standard deviations and a shock in hospitalization as a year-to-year *increase* of one or more standard deviations in the number of nights in hospital.[13] In our data, one standard deviation of the health-satisfaction distribution corresponds to two points on the eleven-point (zero-to-ten) scale for mothers and fathers.[14] Thus, a shock is defined as a drop of four or more points from one year to the following year. For nights spent in hospital, one standard deviation corresponds to six nights during the year for mothers and four nights for fathers; shocks are defined as increases of six or four nights per year for mothers and fathers, respectively.

Based on the two health measures and on combinations of them, we create four alternative shock definitions: (1) a shock in health satisfaction *or* in nights of hospi-

[13]We opt for the more generous definition of a hospitalization shock in order to obtain a reasonable number of shock observations in our sample. If we used the two standard deviation threshold for hospital-defined health shocks as well, we would observe very few cases of maternal (18 cases) and paternal (14 cases) health shocks for children aged 3-6. Our definition of a shock in health satisfaction is similar to and only slightly less stringent than that used by Riphahn (1999), which is a drop by at least five points for older workers. Our results do not substantially change if we use the two standard deviation threshold for hospitalization shocks. These results are available upon request from the authors. Also note that maternal hospitalization shocks imply a duration exceeding the usual hospitalization periods around childbirth. According to Schneider (2008), German childbearing mothers spend on average 2.8 days in hospital.

[14]The definition of a shock necessitates that a parent starts the observation period with a health satisfaction rating that allows for a drop of four points. In our sample, only five percent of parents report health satisfaction scores below the minimum threshold for a shock (satisfaction ≤ 3). Removing these low-health-satisfaction respondents from our sample does not alter our results.

talization; (2) a shock in health satisfaction; (3) a shock in nights of hospitalization; and (4) a shock in both health satisfaction *and* nights of hospitalization.[15]

We assume that shocks are often due to exogenous sources of variation. Ideally, we would also observe the reason for a health shock (e.g. accidents) and only draw on such changes that are least likely to suffer from an omitted-variable bias. However, distinguishing specific health events is not possible since the underlying dataset offers only limited information on symptoms rather than on specific illnesses. Therefore, we use a rather broad health shock definition but additionally control for the initial health of children (at birth) and of their parents (before childbirth). In further specifications, to take general living conditions into account, we also include household characteristics such as household income and parental employment.

Similar shock-based approaches have previously been used in the empirical literature on health effects. For example, Hagan et al. (2009) apply this approach to study health and retirement in Europe. Riphahn (1999) examines impacts of health shocks on income and employment. Both studies define health shocks as standard deviation-based changes in the health variable. As noted above, we use a similar definition for shocks.

Assuming that our observed health shocks are exogenous, conditional on initial health and other covariates, we estimate the following reduced-form specification:

$$Y_i = \alpha + \beta_1 MHS_i^{0-3} + \beta_2 MHS_i^{3-6} + \gamma_1 PHS_i^{0-3} + \gamma_2 PHS_i^{3-6} + X_i\delta_1 + C_i\delta_2 + \epsilon_i \quad (1)$$

Child behavioral skill Y_i is derived from the child's SDQ score at age six. The maternal health shock (MHS) and paternal health shock (PHS) variables distinguish whether the specific parent was subject to a health shock before or after the child is three years of age, as noted by the superscripts 0-3 and 3-6. All of our regressions control for variables that are considered to be related to the children's initial skills endowment: The vector X_i includes child's gender, birth order, week of gestation at birth, birth weight, a second-order polynomial of the age of the mother at the birth (measured in months), and a second-order polynomial of the age of the child (mea-

[15] We observe that roughly 30 percent of mothers who experience a negative change of more than two standard deviations in health satisfaction experienced a corresponding shock, defined by the number of nights spent in hospital, early in children's lives.

sured in months), allowing for a flexible age effect.[16] The vector C_i captures parental education, immigration background, prenatal household income, parental initial health satisfaction, and the initial number of nights of parental hospitalization (all variables were observed one year before the child's birth).[17] Table 1.7 in the appendix lists the control variables along with their means and standard deviations.

The numbers of health shocks observed in our data are provided in Tables 1.2 and 1.3 together with the results for each of the four definitions of a health shock. Generally, maternal health shocks occur more frequently than paternal health shocks in the data.[18] Since there are few paternal health shocks according to our strictest specification (4), we note that the results of this specification are subject to large standard errors.

1.4 Results

Table 1.2 shows the estimated impacts of parental health shocks on children's behavioral skills. For each of the observed outcomes (Panels A to F), we estimate the four specifications according to our definitions of parental health shocks. As indicated above, specification (1) uses the broadest definition of a shock, while specification (4) implies the most restrictive definition. The bottom panel of Table 1.2 provides the respective numbers of health shock observations.

In the following, we first discuss the results for *maternal* health shocks with a focus on statistically significant estimates. Since fewer estimated effects from *paternal*

[16] We use an indicator for firstborn children to take birth order into account. As an alternative to the birth-weight specification, we tested specifications that included a high-birth-weight dummy variable (2.5 kilogram or more). The results are robust if we use this specification.

[17] We control for parental education based on a tertiary education indicator taking the value of one if at least one parent obtained a university degree or a comparable level of (vocational) education. The definition corresponds to a level of 5 or higher according to the UNESCO's International Standard Classification of Education (ISCED 1997), where ISCED level 5 is defined as the "first stage of tertiary education (not leading directly to an advanced research qualification)" (UNESCO, 2006, p. 19). To control for migration background we use an indicator variable taking the value of one if at least one parent holds a non-German nationality. This is true for 16 percent of individuals in our sample, mostly with Turkish (40 percent) and Italian (13 percent) nationality. We also include indicator variables for missing parental information, which take into account that there are single-headed households.

[18] This is also the case in previous studies on parental health effects (e.g. Morefield 2010) and in line with evidence that women in the relevant age group have more episodes of hospitalization than men (cf. Case and Paxson 2005). However, part of the difference reflected in our numbers is due to the fact that there are households headed by single mothers in our sample. According to Table 1.1, about 21 percent of children are growing up without a father. All specifications consider both maternal and paternal health shocks, while we include an indicator variable for households headed by single mothers.

health shocks are statistically significant from zero, these will be addressed in a separate paragraph below.

The regression results for the overall difficulty score suggest that maternal health shocks occurring in the second period of observation – when children are aged 3 to 6 – significantly affect child behavior. The estimates are robust among all specifications and range from 0.4 to 0.9 standard deviations, with the largest effect corresponding to the most stringent measure of a health shock (Panel A of Table 1.2).

Panels B to E of Table 1.2 provide evidence on each of the specific behavioral dimensions. For specifications (2) and (4), maternal health shocks occurring when children are aged 0-3 are negatively related to the emotional symptom score. The estimated impact varies from 0.2 (specification 2) to 0.4 standard deviations (specification 4). However, the estimate is not robust when shocks are defined by hospitalizations (specifications 1 and 3). In contrast, second period maternal health shocks more consistently exert a negative impact of about one half of a standard deviation: Again, we see the strongest impact when looking at the most restrictive health-shock definition (0.6 standard deviations).

For second period maternal health shocks, we also observe negative impacts of about 0.2 standard deviations on the child hyperactivity score (Panel C). However, the respective estimates are not statistically significant from zero in specifications (3) and (4), due to low power from a smaller number of observed health shocks. Similarly, second period maternal health shocks seem to be related to a higher incidence of children's conduct problems (Panel D). The size of the estimated effect is about 0.2 standard deviations (not statistically significant in specification (3), with a larger impact of 0.9 standard deviations for specification (4).

Maternal health shocks are not robustly related to child peer relationship problems (Panel E). Although all of the point estimates suggest a higher incidence of peer relationship problems for second period maternal health shocks, this result is only statistically significant in specification (2), where the estimate amounts to 0.2 standard deviations. And, in contrast, the signs of the point estimates for earlier maternal health shocks are not consistently positive or negative in specifications (1) to (4), with a positive and statistically significant estimated impact of about 0.2 standard deviations when based on hospitalization shocks (specifications 1 and 3). We do not observe any significant impact of parental health shocks on children's pro-social behavior (Panel F).

Table 1.2: Impact of health shocks on child behavior (age 6)

Specifications		(1) any shock	(2) shock in health satisfaction (2 s.d.)	(3) shock in hospitalization (1 s.d.)	(4) shock in satisfaction & hospitalization
(A) Overall difficulty score, SEB (R)	Mother	0.01	-0.17	0.06	-0.23
	(age 0-3)	(0.09)	(0.12)	(0.11)	(0.24)
	Father	0.06	0.45***	-0.15	0.23
	(age 0-3)	(0.13)	(0.17)	(0.15)	(0.43)
	Mother	-0.41***	-0.42***	-0.36**	-0.95*
	(age 3-6)	(0.10)	(0.13)	(0.15)	(0.51)
	Father	-0.01	0.12	-0.23	-0.19
	(age 3-6)	(0.12)	(0.13)	(0.17)	(0.32)
(B) Emotional Symptoms (R)	Mother	-0.04	-0.21*	0.01	-0.37*
	(age 0-3)	(0.09)	(0.13)	(0.11)	(0.23)
	Father	-0.00	0.19	-0.06	0.46
	(age 0-3)	(0.13)	(0.19)	(0.16)	(0.29)
	Mother	-0.47***	-0.40***	-0.47***	-0.62**
	(age 3-6)	(0.11)	(0.13)	(0.15)	(0.29)
	Father	0.11	-0.25*	-0.07	0.27
	(age 3-6)	(0.12)	(0.13)	(0.17)	(0.21)
(C) Hyperactivity (R)	Mother	-0.11	-0.18	-0.07	-0.15
	(age 0-3)	(0.09)	(0.12)	(0.11)	(0.25)
	Father	0.07	0.38**	-0.13	-0.24
	(age 0-3)	(0.12)	(0.16)	(0.16)	(0.50)
	Mother	-0.21**	-0.26**	-0.17	-0.82
	(age 3-6)	(0.10)	(0.12)	(0.16)	(0.51)
	Father	-0.07	-0.02	-0.20	-0.48
	(age 3-6)	(0.12)	(0.14)	(0.16)	(0.36)
Observations		639	639	639	639
# maternal shocks (age 0-3)		183	78	126	21
# paternal shocks (age 0-3)		74	30	48	4
# maternal shocks (age 3-6)		138	85	61	8
# paternal shocks (age 3-6)		86	54	39	7

Continued on next page...

... table 1.2 continued

Specifications		(1) any shock	(2) shock in health satisfaction (2 s.d.)	(3) shock in hospitalization (1 s.d.)	(4) shock in satisfaction & hospitalization
(D) Conduct Problems (R)	Mother	0.04	-0.05	0.06	0.06
	(age 0-3)	(0.09)	(0.13)	(0.11)	(0.22)
	Father	0.05	0.52***	-0.20	0.52
	(age 0-3)	(0.13)	(0.15)	(0.15)	(0.36)
	Mother	-0.20**	-0.22*	-0.22	-0.85**
	(age 3-6)	(0.10)	(0.12)	(0.15)	(0.43)
	Father	-0.06	0.03	-0.20	-0.35
	(age 3-6)	(0.12)	(0.15)	(0.18)	(0.51)
(E) Peer relationship problems (R)	Mother	-0.19**	-0.00	0.21*	-0.16
	(age 0-3)	(0.10)	(0.14)	(0.11)	(0.27)
	Father	0.06	0.06	0.03	-0.42
	(age 0-3)	(0.13)	(0.22)	(0.15)	(0.58)
	Mother	-0.16	-0.23**	-0.02	-0.12
	(age 3-6)	(0.10)	(0.12)	(0.14)	(0.36)
	Father	-0.04	0.04	-0.16	0.03
	(age 3-6)	(0.13)	(0.15)	(0.18)	(0.23)
(F) Pro-social behavior, PBS	Mother	0.01	-0.08	0.03	-0.10
	(age 0-3)	(0.09)	(0.12)	(0.10)	(0.22)
	Father	-0.09	0.06	-0.21	-0.56
	(age 0-3)	(0.12)	(0.17)	(0.15)	(0.57)
	Mother	0.01	-0.08	0.09	-0.20
	(age 3-6)	(0.10)	(0.12)	(0.14)	(0.38)
	Father	0.14	0.11	0.17	0.33
	(age 3-6)	(0.11)	(0.14)	(0.14)	(0.29)
Observations		639	639	639	639
# maternal shocks (age 0-3)		183	78	126	21
# paternal shocks (age 0-3)		74	30	48	4
# maternal shocks (age 3-6)		138	85	61	8
# paternal shocks (age 3-6)		86	54	39	7

Notes: Coefficients (robust standard errors) from regressions using the SOEP, version 28. Regressions control for parental initial health satisfaction and hospitalization, education and immigrant background, household income before giving birth, children's gender, birth order, gestation week of pregnancy at birth date, birth weight, a second-order polynomial of the age of the mother at childbirth and a second-order polynomial of the children's age at the time of observation. Outcome variables are standardized test scores (z-score). Higher scores imply more favorable behavioral skills (R = reverse scale). * Statistically significant at the ten percent level of significance, ** at the five percent level and *** at the one percent level. s.d. = standard deviation.

For all of the outcomes, we fail to find evidence for *negative* impacts of paternal health shocks. However, for specification (2), we obtain statistically significant positive coefficients of early paternal health shocks in panels A, C (conduct problems) and D (hyperactivity), suggesting improvement after a parental health shock. We do not observe such a positive impact related to shocks defined by changes in hospitalizations.[19] However, the positive point estimates are rather high, suggesting that there might be a direct positive effect of paternal health problems (not related to hospitalization) on children's hyperactivity (0.4 standard deviations) and conduct problems (0.5 standard deviations). One explanation for the positive estimated impacts may be that sick fathers shift market time towards time with children, contributing positively to children's skill development.[20] The increase in paternal time investments may be more beneficial for children's skill development than a reduction in goods investments, due to lost income. Moreover, the reduction in goods investments may be mitigated in Germany, where sick pay regulations are relatively generous. To this end, a detailed time use analysis is beyond the scope and feasibility of this paper.

In sum, the general pattern observed from Table 1.2 is that *paternal* health shocks do not exert consistent impacts on behavioral skills, while we find negative and often significant impacts of *maternal* health measures. This is especially true for maternal health shocks observed after children are aged 3. The latter finding is in line with evidence from the US (Morefield 2010) who finds that the onset of parental health conditions is related to significant increases in children's problem behavior at ages 5-9 (but not at younger or older observed age categories). This finding suggests that non-cognitive skills are more affected by poor parental health during the stage of childhood around school entry age.[21] This is also in line with evidence from neuroscience

[19]One explanation for the different findings related to optional health shock definitions may be that shocks in health satisfactions relate to different kind of health conditions than shocks in hospitalization. For example, it may be that shocks in health satisfaction are more often due to mental health problems while hospitalization shocks are somewhat more often due to physical health limitations. To this end, our data does not allow distinguishing between mental and physical health conditions to further investigate this issue. We are grateful to an anonymous referee for suggesting this potential explanation.

[20]A simple time-allocation model suggests this type of trade-off if the wage of the father is reduced and the substitution effect dominates the income effect.

[21]Based on our results, we cannot rule out that a contemporaneous negative effect from parental health shocks earlier in children's life already fades out when children are observed (at age 6). The finding in Morefield (2010) based on a longer period of observation hints to a non-significant impact of early health shocks and a permanent effect of the parental health shock for children around school-entry age, but in our data, we do neither observe a longer period of time nor the behavioral outcome for three year-olds. Therefore, we cannot directly test for dynamic effects.

demonstrating that socio-emotional skills are developed throughout different critical and sensitive periods during childhood and adolescence (e.g. Tonks et al. 2009).[22]

In a next step, we conduct robustness checks in order to challenge our identifying assumption. If parental health shocks identify causal effects, we would expect that parental health shocks would affect future child skills outcomes while they are not correlated to past child skills outcomes. In other words, if the same parental health shocks that occur when the child is age six affect child outcomes at age three then this will challenge the validity of our empirical approach. We conduct the robustness checks using the Vineland Adaptive Behavior Scale (VAB). In our data, the VAB is measured for children at age three, whereas the SDQ outcome is measured at age six. We observe parental health shock information throughout the first six years of children's lives. Thus, we are able to relate children's VAB to parental health shocks prior to and after the VAB is measured at age three. First, we regress the Vineland score on an indicator of parental health shocks prior to child assessment and an indicator for a shock occurring in the future.[23] We control for the same set of control variables that are included in our main regressions. As a second check, we regress adaptive behavior solely on future parental health shocks and control variables ("placebo regressions").

[22] Tonks et al. (2009) summarize that "as the demands of the social environments increase with development, emotion-recognition abilities undergo periods of development in response" (ibid., page 12). For example for social understanding they provide evidence that specific skills are developed throughout the childhood years (e.g. the ability to comprehend misconceiving situations is developed around age 4). The authors note that these development periods are related to development stages of the prefrontal cortex in childhood and adolescence.

[23] Our results may still be biased if (unobserved) events that affect both parental health and child outcomes systematically happen around the time of the reported health shocks. However, in a further robustness check we control for indicators of such events (see the discussion of Table 1.4) and find robust results.

Table 1.3: Impact of health shocks on adaptive behavior and robustness checks (age 3)

Specifications		(1) any shock	(2) shock in health satisfaction (2 s.d.)	(3) shock in hospitalization (1 s.d.)	(4) shock in satisfaction & hospitalization
Effects of past parental health shocks at age 3					
(A) Past health shocks	Mother	-0.18*	-0.16	-0.16	-0.21
	(age 0-3)	(0.09)	(0.13)	(0.10)	(0.28)
	Father	0.13	-0.03	0.21	-0.21
	(age 0-3)	(0.10)	(0.14)	(0.13)	(0.52)
Robustness check 1: Effects of past and future parental health shocks at age 3					
(B) Past health shocks	Mother	-0.18**	-0.16	-0.14	-0.22
	(age 0-3)	(0.09)	(0.13)	(0.10)	(0.28)
	Father	0.13	-0.04	0.23*	-0.20
	(age 0-3)	(0.10)	(0.14)	(0.13)	(0.51)
Future health shocks	Mother	0.05	0.14	-0.05	0.27
	(age 3-6)	(0.10)	(0.11)	(0.15)	(0.33)
	Father	-0.01	0.11	-0.14	0.09
	(age 3-6)	(0.10)	(0.10)	(0.16)	(0.26)
Robustness check 2: Effects of future health shocks at age 3 (Placebo regressions)					
(C) Future health shocks	Mother	0.04	0.14	-0.06	0.25
	(age 3-6)	(0.10)	(0.10)	(0.15)	(0.37)
	Father	-0.01	0.12	-0.14	0.09
	(age 3-6)	(0.10)	(0.10)	(0.16)	(0.26)
Observations		634	634	634	634
# maternal shocks (age 0-3)		180	75	128	23
# paternal shocks (age 0-3)		70	29	44	3
# maternal shocks (age 3-6)		143	87	63	7
# paternal shocks (age 3-6)		86	53	39	6

Notes: Coefficients (robust standard errors) from regressions using the SOEP, version 28. Regressions control for parental initial health satisfaction and hospitalization, education and immigrant background, household income before birth, childrens gender, birth order, gestation week of pregnancy at birth date, birth weight, a second-order polynomial of the age of the mother at the time of childbirth and a second-order polynomial of the childrens age at the time of observation. Outcome variables are standardized test scores (z-score). Higher scores imply more favorable skills. * Statistically significant at the ten percent level of significance, ** at the five percent level and *** at the one percent level. s.d. = standard deviation.

Table 1.3 shows that past maternal health shocks negatively impact child adaptive behavior at age three. The point estimates are consistently negative with statistical significance in specification (1). Even though we lack significance in the other specifications, the magnitudes of the estimated coefficients are robust and amount to 0.2 standard deviations. Evidence on future health shocks is provided in the bottom panels of Table 1.3. The negative impact of past maternal health shocks on adaptive behavior is maintained when future health shocks are included. The results consistently imply that future health shocks are *not* significantly related to children's adaptive behavior at age 3. This is true if future and past health shocks are included in the regressions. It also holds if only future health shocks are included in the regressions. None of the health shock coefficients in the placebo regressions are statistically significant, and the point estimates for maternal health shocks are positive rather than negative. Again, the large coefficient in specification (4) is statistically indistinguishable from zero due to a low number of health-shock observations.

Table 1.4 repeats our main regression results and adds a new set of control variables to the specifications. These variables include indicators for maternal or paternal job losses, parental separation or divorce, and the average household net-income in each year of observation. In sum, these variables reflect the current (economic) situation of the household. They may also represent events that could both trigger or result from health shocks, and potentially exert a direct effect on children's behavioral skills.

Table 1.4 shows that including household event variables does not qualitatively change the estimated coefficients of parental health shocks.[24] Note that the additional control variables specifically contain household income. Therefore, one interpretation of these robust findings is that reduced financial resources do not seem to be the main driver of the observed effects. Consequently, the effects may be due to limited quality or quantity of parental time. However, given the limitations of the available data, a direct examination of these mechanisms reaches beyond the scope of our paper.[25]

[24]In addition, our results are robust if we include additional indicator variables for positive shocks in health satisfaction. However, including only these positive shock measures yields "effects" of about 50 percent of the statistically significant estimates obtained from the original regressions. This finding is probably due to parents being affected from a negative shock, followed by a period of recovery.

[25]Due to our limited sample size, we do not differentiate the results between boys and girls. However, regressing gender-specific outcomes on aggregated parental health shock indicators suggests that impacts are more pronounced for boys (not shown here). This finding is consistent with Johnson and Reynolds (2013) and Morefield (2010).

Table 1.4: Robustness checks including household control variables

Specifications		(1) any shock	(2) shock in health satisfaction (2 s.d.)	(3) shock in hospitalization (1 s.d.)	(4) shock in satisfaction & hospitalization
(A) SEB (R)	Mother	-0.41***	-0.42***	-0.36**	-0.95*
	(age 3-6)	(0.10)	(0.13)	(0.15)	(0.51)
(A) + additional controls	Mother	-0.40***	-0.40***	-0.37**	-0.95*
	(age 3-6)	(0.11)	(0.13)	(0.15)	(0.51)
(B) Emotional Symptoms (R)	Mother	-0.47***	-0.40***	-0.47***	-0.62**
	(age 3-6)	(0.11)	(0.13)	(0.15)	(0.29)
(B) + additional controls	Mother	-0.45***	-0.38***	-0.47***	-0.65**
	(age 3-6)	(0.11)	(0.12)	(0.16)	(0.31)
(C) Hyperactivity (R)	Mother	-0.21**	-0.26**	-0.17	-0.82
	(age 3-6)	(0.10)	(0.12)	(0.16)	(0.51)
(C) + additional controls	Mother	-0.22**	-0.23*	-0.19	-0.83*
	(age 3-6)	(0.10)	(0.12)	(0.15)	(0.50)
(D) Conduct Problems (R)	Mother	-0.20**	-0.22*	-0.22	-0.85**
	(age 3-6)	(0.10)	(0.12)	(0.15)	(0.43)
(D) + additional controls	Mother	-0.20*	-0.21*	-0.22	-0.82**
	(age 3-6)	(0.10)	(0.12)	(0.15)	(0.41)
Observations		639	639	639	639
# maternal shocks (age 0-3)		183	78	126	21
# paternal shocks (age 0-3)		74	30	48	4
# maternal shocks (age 3-6)		138	85	61	8
# paternal shocks (age 3-6)		86	54	39	7

Notes: Coefficients (robust standard errors) from regressions using the SOEP, version 28. Regressions control for parental initial health satisfaction and hospitalization, education and immigrant background, household income before birth, children's gender, birth order, gestation week of pregnancy at birth date, birth weight, a second-order polynomial of the age of the mother at the time of childbirth and a second-order polynomial of the children's age at the time of observation. '+ additional controls' include maternal or paternal job losses, parental separation or divorce and the average monthly family income in the period of observation. Outcome variables are standardized test scores (z-score). Higher scores imply more favorable skills. * Statistically significant at the ten percent level of significance, ** at the five percent level and *** at the one percent level. s.d. = standard deviation.

1.5 Conclusion

In line with previous studies, our paper highlights the importance of parental investments in children's skill formation. One interpretation of our findings is that an involuntary change in parental (maternal) investment yields negative impacts on children's behavioral skill development. Specifically, maternal health shocks that occur for children aged 3-6 are found to negatively affect the emotional symptoms, conduct problems, and hyperactivity observed in six-year-old children. These negative effects range up to 0.9 standard deviations, with more negative point estimates found as we use more stringent definitions of health shocks. We do not observe negative effects of paternal health shocks or of shocks that occur earlier in children's lives. Our results hold throughout several robustness checks.

The general findings of negative parental health effects on behavioral outcomes measured around school entry age are in line with the above-mentioned study for the US, Morefield (2010). However, Morefield (2010) also finds that *paternal* health plays a more important role in shaping children's behavioral skills in the US. Because of different samples and estimation strategies, the findings are not directly comparable to ours (cf. footnote 6). As discussed in the introduction of our study, one explanation for different impacts of parental health in both countries may be institutional differences. More generous parental benefit and sick pay regulations in Germany facilitate the role of German mothers as predominant caregivers of their children, while the "breadwinner's" health is less important with respect to sustaining family income. These stylized institutional differences point to a higher importance of consistent maternal health in Germany as compared to the US.

While we observe negative impacts of maternal health on children's behavioral skills at age six, we cannot address long-run effects. From a theoretical point of view, the nature of the technology of skill formation suggests that a drop in skills in one period negatively affects skill development in later periods (Cunha et al. 2006). Still, lack of empirical evidence on long-run effects of parental health provides scope for future research (depending on the availability of data). If our results translated to long-run impacts in forming children's skills, this would point to the importance of measures to support sick parents. To this end, additional support by external caregivers and home-visit programs may be effective.[26] Evaluation of existing programs also provides scope for further research but is restricted by the availability of appropriate data.

[26] According to the German Social Security Act (*Sozialgesetzbuch*) sick parents can obtain domestic help (*Haushaltshilfen*), which is organized by the parents' health insurance company.

Appendix

Table 1.5: Items of the Strengths and Difficulties Questionnaire (SDQ)

Hyperactivity:
"Child is agitated, overactive, cannot sit still"; "child is fidgety"; "child is easy to distract, cannot concentrate"; and "*child finishes tasks,can concentrate"*.[27]

Emotional Symptoms:
"Child is often unlucky or sad, cries often"; "child is nervous in new situations, cramps"; and "child is very anxious, frightens easily".

Conduct Problems:
"Child tends to have a fit of rage, is explosive"; and "child argues often with other children, bullies them".

Peer Problems:
"Child is a maverick"; "*child is popular*"; "child is often fooled by others, is bullied"; and "child gets along with adults better than with other children".

Pro-social behavior:
"*Is thoughtful*"; "*child likes to share with others*"; "*child is helpful, if others are hurt, ill or sad*"; and "*child helps others willingly*".

Source: According to Goodman (1997), translations are the English labels of the SOEP items.

[27] *Italic SDQ items correspond to a positive characteristic.* The scales of all other items are reversed to yield a positive meaning when aggregating scores.

Table 1.6: Items of the Vineland Adaptive Behavior Scale (VAB), SOEP-version

Talking:
“Understands brief instructions, such as ‘go get your shoes’ ”; “Forms sentences with at least two words”; “Speaks in full sentences (with four or more words)”; “Listens attentively to a story for five minutes or longer”; “Passes on simple messages such as ‘dinner is ready’ ”.

Everyday skills:
“Uses a spoon to eat without assistance and without dripping it”; “Blows his/her nose without assistance”; “Uses the toilet to do ‘number two’ ”; “Puts on pants and underpants the right way around”; “Brushes his/her teeth without assistance”.

Movement:
“Walks forwards down the stairs”; “Opens doors with the door handle”; “Climbs up playground climbing equipment and other high playground structures”; “Cuts paper with scissors”; “Paints/draws recognizable shapes on paper”.

Social relationships:
“Calls familiar people by name; for example, says ‘mommy’ and ‘daddy’ or uses the father’s first name”; “Participates in games with other children”; “Gets involved in role-playing games ‘playing pretend’)”; “Shows a special liking for particular playmates or friends”; “Calls his/her own feelings by name, e.g. ‘sad’, ‘happy’, ‘scared’ ”.

Source: Translations are the English labels of the SOEP items.

Table 1.7: Means (standard deviations) of control variables included in the main regression analysis

	Socio-emotional development sample	Adaptive behavior sample
Tertiary education of parents indicator	0.59	0.59
(observed at birth)	(0.49)	(0.49)
Indicator for missing parental education	0.08	0.09
	(0.27)	(0.28)
Parental migration background indicator	0.16	0.16
(observed at birth)	(0.35)	(0.37)
Indicator for missing parental migration	0.08	0.09
background	(0.28)	(0.29)
Maternal initial health satisfaction	7.58	7.56
(observed before birth)	(1.89)	(1.90)
Paternal initial health satisfaction	7.52	7.51
(observed before birth)	(1.90)	(1.88)
Maternal initial hospitalization nights	1.15	1.21
(observed before birth)	(4.38)	(4.48)
Paternal initial hospitalization nights	0.80	0.67
(observed before birth)	(4.82)	(4.00)
Monthly household income	2527.68	2528.36
(CPI adjusted, observed before birth)	(1261.70)	(1267.80)
Indicator for missing household income	0.17	0.17
(not observed before birth)	(0.38)	(0.38)
Gender: male indicator	0.48	0.47
	(0.50)	(0.50)
Age of child	69.06	69.11
(in months, last measurement point)	(3.80)	(3.82)
Birth order: first born indicator	0.45	0.45
	(0.50)	(0.50)
Age of mother at birth	30.78	30.73
	(5.27)	(5.30)
Week of pregnancy at childbirth	39.12	39.09
	(2.32)	(2.37)
Birth weight of child (in g)	3339.88	3333.36
	(573.96)	(583.29)
# Observations	639	634

Source: Mother-and-child data of the German Socio-Economic Panel Study (SOEP), version 28. Own calculations.

Chapter 2

The Impact of Lengthening the School Day on Substance Abuse and Crime: Evidence from a German high school reform

2.1 Introduction

Cannabis is the most consumed illegal drug in Germany and ranks third after legal substance alcohol and nicotine.[1] Furthermore, cannabis is the most prominent illegal drug among the youth in Germany and is considered responsible for two-thirds of drug-related crimes among youth. Regardless of recent efforts to decriminalize the consumption of cannabis, it is an undisputed aim to prevent children and adolescents from consuming any form of drug. To educate students about drugs and their risks, the school curriculum includes lessons on substance abuse prevention and is supported by no-drug campaigns in Germany.

This study analyzes the effect of lengthening a school day in the academic high school track on illegal substance abuse and criminal behavior. Following Germany's school reform, the final year of high school was eliminated, and the instructional hours

[1]See Drogenbeauftragte (2013) for an overview of drug consumption in Germany. Alcohol spirits and nicotine are forbidden substances for adolescents aged below 18 years. Beer and wine are prohibited for those aged below 16 years.

were shifted to lower grades. This increase in school hours reduces opportunities that students have to obtain and consume drugs. I estimate the impact of an increase in instructional hours on the crime rates of the age groups affected by the reform. Affected students receive the same total number of instructional hours during graduation, which forces them to spend more time at school in lower grade levels. This shift mainly affected the middle grades of the high school period, whereas an increase in instructional hours during the last years before graduation was moderate, because of an already dense curriculum at these grade levels.

The link between education and crime has been extensively studied in the past (Ehrlich, 1975; Jacob and Lefgren, 2003; Lochner and Moretti, 2004). Most studies in this area rely on exogenous variation via reforms in the institutional framework to assess the impact of education on high-risk youth groups with respect to crime. The economic literature distinguishes between two channels through which education can impact crime. The first channel works through an investment in education, which increases the opportunity costs of committing crimes. The second channel works through incapacitation in school or education and can be explained by the fact that the time spent at school cannot be used to commit crimes. This incapacitation effect does not necessarily depend on the quality of education. Kline (2012) finds that simple curfews, as a form of incapacitation, are effective at reducing both violent and property crimes of juveniles.

This study differs from previous literature in that I can estimate the effect of schooling on crimes committed by high-performing students who should be a relatively low-risk youth group with respect to delinquent behavior. Students in an academic high school track intend to pursue a school career beyond the minimum dropout age and thus are not affected by changing it.

To estimate the causal impact of the high school reform on crime rates, the underlying analysis applies a difference-in-difference strategy, which uses variation over time between the *Länder* (German Federal States) and the affected age groups.[2] The results suggest that the increase in instructional hours at lower grades slightly decreases overall crime rates of the affected age groups. However, the drop in crime rates is mostly driven by declining violent and drug-related crimes. Furthermore, in-depth analysis reveals that drug-related crimes decline as a consequence of decreasing arrest rates of cannabis users in the affected age groups.

[2]Please see Chevalier and Marie (2016) for a description of the research method and Chevalier and Marie (2013) for another application of the crime data.

A common approach to study the exogenous effects of education on crime is to monitor changes in the minimum dropout age (cf. Anderson 2014, Machin et al. 2011, and Brilli and Tonello 2015). This identification strategy ensures that potential offenders are affected by the reform because early school dropouts show higher offender rates than classmates who remain at school. Given that these students have lower opportunity costs in education than high-performing students, one can suspect a stronger effect of education on crime for low-performing students. However, to the best of my knowledge, the evidence of the educational effect on crime for high-performing students who are hardly affected by minimum dropout age regulations is scarce, especially with regard to drug-related crimes.

Jacob and Lefgren (2003) present evidence on the contemporaneous effect of schooling on crime in the US using in-service days of teachers as a source of exogenous variation.[3] Their results suggest that students who are incapacitated at school have relatively fewer possibilities to prepare for or commit criminal activities, at least in the case of property crime. However, incapacitation of students in school increases violent criminal behavior in the US. Similar evidence is provided by Luallen (2006) who explores the effect of teacher strike days (reduced classroom teaching time) on criminal activities. These measured incapacitation effects are in general comparable with the reform's effect of increased instructional hours discussed in this study, however, they cannot differentiate between specific secondary school types and their students.

Deming (2011) uses lotteries for attending first-choice schools to estimate the impact of education on crime rates seven years after graduation. He shows that 'winning' students benefit from a higher school quality through better qualified teachers and from less crime prone peer groups. Furthermore, he finds that 'winning' students have lower crime and incarceration rates during and after their school careers. He notes that especially in the age group of middle-school children, the peer group effect has a strong negative influence on violent crime. However, he admits that the 'winning' students are drawn from a population with a low social and economic background. These results are in line with Becker (1968) and can be explained by the increased opportunity costs of crime due to potentially higher earnings in the legitimate sector.

Åslund et al. (2015) study the effect of education on crime with a Swedish reform of the vocational upper secondary education, extending the curriculum from two to three years. This reform targets students who are a high-risk group with respect to criminal

[3] Teachers have to deal with administrative duties on these in-service days, while students do not have to attend school. Jacob and Lefgren (2003) show that these in-service days are more exogenous than weekends or national holidays.

behavior. They find a negative effect on property crime but not on violent crime of the additional school year, which can be explained by an incapacitation effect too. In comparison to this study with a focus on high-performing students, the underlying reform of their study affected mostly low-performing students. Berthelon and Kruger's (2011) study relates most closely to the present research. They use a school reform in Chile that lengthened the school day for public and publicly funded private schools. The lengthened school day was found to reduce the likelihood of teenage pregnancy and decrease juvenile property and violent crimes.

To my knowledge, the current study is the first that examines an educational reform impact with a focus on substance abuse. The staggered implementation of the reform over different *Länder*, grade levels, and years serves as a source of the exogenous variation of schooling via instructional hours. The design of the reform allows for an evaluation approach, which explores differences over time, age cohorts, and between *Länder*.

The present study also provides evidence from a regional student drug survey taken in the *Land* of Hamburg that supports the findings from the police crime data. It shows that the reduction of the cannabis users after the reform was driven by a decreasing consumption by the reform-affected students and rejects the hypothesis of drug consumption shifted into school hours.

The remainder of the paper is organized as follows. Section 2.2 provides a literature review and the institutional framework. Section 2.3 describes the datasets. Section 2.4 explains the identification and estimation strategy. Section 2.5 presents the results and Section 2.6 concludes the study.

2.2 Literature Review and Institutional Framework

2.2.1 Literature Review

Several studies use the German high school reform implementation to evaluate the impact of the eliminated school year on educational achievements and other outcomes.[4] Huebener and Marcus (2014) find that the main goal of the reform, reducing graduation age, was achieved. But, the rate of grade repetition shortly before graduation doubled after the reform. It is not clear whether the increase in grade repetition rates is a long-term effect or is driven by frictions during the implementation process. Büttner and Thomsen (2015) find negative effects of the reform on grades in math in one *Land*.

A similar reform occurred in the U.S., which reduced the days per week that students spend in school. Anderson and Walker (2015) analyze the effect of a shift from a five- to four-day school week on student achievement in rural areas. To ensure the minimum state-mandated requirements, these schools needed to increase instructional hours per day. Anderson and Walker find positive rather than negative effects of the lengthened school day on math and reading test scores.

Dahmann and Anger (2014) discover that after the German high school reform students are slightly more extroverted and less emotionally stable compared with those from non-reformed high schools. A psychological survey by Milde-Busch et al. (2010) does not find any difference in headache and other stress measures after the reform, but results do show that students with a lengthened school day declare fewer hours of spare time.

2.2.2 German Education System

Education policy is not centralized in Germany, but it is one of the main competences of German *Land's* jurisdictions, and federal responsibilities are limited. *Länder* can reform their education systems independently of each other. However, a voluntary assembly of *Länder* ministers of education coordinates reforms and ensures a comparable set of standards.

In Germany, students begin schooling close to their 6th birthday at enrollment, in primary school. Given this average age of students within primary school, students are,

[4]Cf. Kühn et al. 2013, Meyer et al. (2015).

roughly, 10 years old when tracking occurs in secondary school (high school).[5] Based on students' competences and preferences, teachers or parents decide which track the students should attend in their secondary school education.[6]

The '*Gymnasium*' is the highest academically focused secondary school track and covers grade levels 5-12 (i.e., 5-13 before the reform).[7] Graduates from these schools receive a general qualification of university entrance and can study at a university or polytechnic tertiary teaching institution without any further training.

The '*Realschule*' track offers a less academic curriculum for secondary school covering grades (5-10), and prepares students for an apprenticeship, typically leading to white-collar jobs. The '*Hauptschule*' track has the least academic curriculum, ending after grade 9, and prepares students for an apprenticeship that will lead to trade or the industrial sector. However, the tracking between Realschule and Hauptschule is less well-defined in some *Länder*, due to comprehensive schools with a curriculum that is more independent from the track.[8]

The *Gymnasium* track covers the majority of students in most *Länder* and was exclusively subject to the high school reform. Similar vocational grammar schools with an equivalent university entrance diploma kept the old curriculum.[9] The 2001 *Gymnasium* track accommodates approximately 30% of a student's age cohort. However, this share has increased over the last 10 years, to approximately 40% of a student's age cohort in 2012.[10]

[5]Three out of sixteen *Länder* track their students after the 6$^{\text{th}}$ grade. Furthermore, some schools offer a curriculum for up to three tracks within one school institution.

[6]Whether parents or teachers decide the optimal secondary school type depends on the legal framework in each *Land*. For a detailed description of tracking in Germany, see Dustmann (2004).

[7]Students who start their secondary school period at other school types can switch to a *Gymnasium* generally on the completion of each school year if their academic performance is high, however, students usually switch tracks after finishing their current school and attend a *Gymnasium* at corresponding later levels.

[8]Beyond the traditional three-track system, Germany offers comprehensive schools *Gesamtschulen* and special schools for children who are physically or mentally challenged (*Förderschulen*). There are also the so-called 'Waldorf schools' that focus on teaching with an anthroposophical approach.

[9]Students in vocational grammar schools can receive university entrance qualification even after graduation; however, the high school phase of vocational schools remained until the 13$^{\text{th}}$ grade and was not affected by the reform.

[10]Based on the Annual Report of general education from the Federal Statistical Office of Germany, 2001-2012.

2.2.3 G8 Reform

The object of the reform, which eliminated the last high school year, was to support students with an earlier entry into the university (or job market) in accordance with international education systems. Hence, the years of secondary education in the *Gymnasium* decreased from nine years (*G9*) to eight years (*G8*).[11]

After the reform, the total number of instructional hours remained constant, however, as the length of the school day increased at the lower grade levels.[12] However, the workload at the first two grades in the *Gymnasium* (grades 5 and 6) was rarely increased to avoid extra burden during the transition from primary to secondary school.[13] And, the already high number of instructional hours at the final stage of the *Gymnasium* (grades 11 and 12) prevent any additional increase of instructional hours at prior to graduation. As a result, the 'shifted' instructional hours of the old 13$^{\text{th}}$ grade were mainly distributed among grades 7 to 10, which increased the instructional hours per day by up to 20%. Many schools switched from a half-day to a full-day program to deal with the reform.

The implementation of the *G8* reform was staggered over the *Länder* and with different grade levels affected first. Table 2.1 provides an overview of the implementation in the German *Länder*. Two East German *Länder*, Saxony and Thuringia, had not changed the high school length after reunification, and were already operating under the *G8* policy. However, the other East German *Länder* adopted the West German *G9* regime after reunification in 1990. The first West German *Land* to introduce the shorter high school system was Saarland, altering the 5^{th} grade in the 2001/2002 school year. The first graduates of this reform finished school in the double *G8* and *G9* graduation year, 2009. However, Saxony-Anhalt was the first *Land* with a double graduation cohort in 2007. As the *G8* reform in the 2003/2004 school year affected the 9^{th} grade and below, the 9^{th} and 10^{th} grades of the 2003/2004 school year graduated together in 2007.[14]

[11] *G9* refers to the old school regime and *G8* refers to the new reformed school regime.

[12] According to the Kultusminsiterkonferenz (KMK, 2013), the average hours per week in the *Gymnasium* increased from 29.44 to 33.13.

[13] Some *Länder* foster easier transitions between the different school tracks at the entrance to the secondary school phase. In the so-called 'Orientierungsstufe' or 'Förderstufe which covers the 5$^{\text{th}}$ and 6$^{\text{th}}$ grade, student tracking is less strict.

[14] For further information regarding the relative short implementation phase in Saxony-Anhalt and the impact on educational outcomes, see Büttner and Thomsen (2015).

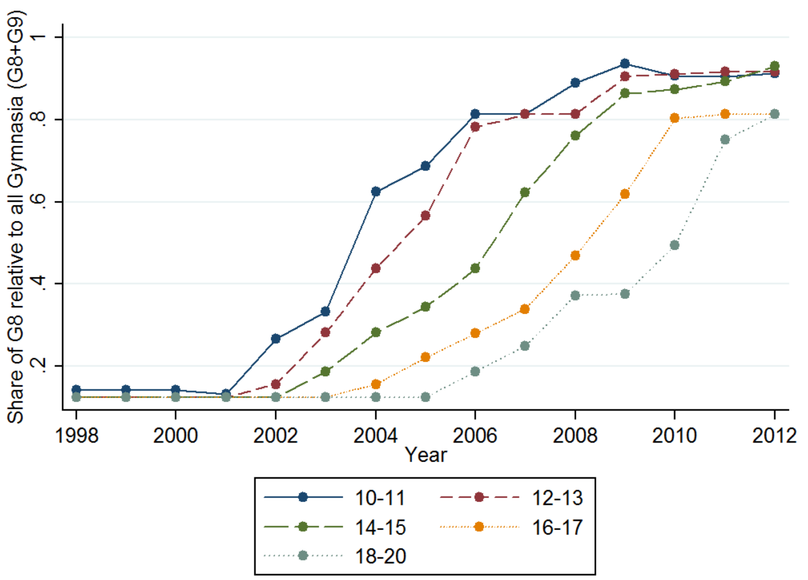
Table 2.1: Implementation timetable of the *G8* reform

Federal Land	Introduction *G8* (School Year)	starting Grade(s)	Double Graduation Year
Baden-Wuerttemberg	2004/2005	Grade 5	2012
Bavaria	2004/2005	Grades 5 & 6	2011
Berlin	2006/2007	Grade 7	2012
Brandenburg	2006/2007	Grade 7	2012
Bremen	2004/2005	Grade 5	2012
Hamburg	2002/2003	Grade 5	2010
	2004/2005: ca. 10% of schools	Grade 5	2012
Hesse	2005/2006: ca. 60% of schools	Grade 5	2013
	2006/2007: ca. 30% of schools	Grade 5	2014
Lower Saxony	2004/2005	Grades 5 & 6	2011
Mecklenburg-West Pomerania	2004/2005	Grades 5 to 9	2008
Northrine-Westphalia	2005/2006	Grade 5	2013
Rhineland-Palatinate	Only pilot schools 2008/2009	Grade 5	—
Saarland	2001/2002	Grade 5	2009
Saxony	since 1949	—	—
Saxony-Anhalt	2003/2004	Grades 5 to 9	2007
Schleswig-Holstein	2008/2009	Grade 5	2016
Thuringia	since 1949	—	—

Source: Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK).
Notes: Four Länder *are partially withdrawing the reform again with 'pilot' schools of the old regime. Lower Saxony will take back the reform completely in the school year 2015/2016.*

Figure 2.1 plots the actual share of reform affected *Gymnasium* students relative to all *Gymnasium* students across the *Länder* for the years 1998 to 2012 by two- and three-year age groups. The staggered implementation starts around 2002. However, due to the *Länder* Saxony and Thuringia, which had the *G8* regime since 1949, a small fraction of *G8 Gymnasia* was already present prior to 2002.

Figure 2.1: Relative share of students in the *G8* track relative to all *Gymnasium* students (*G8* and *G9*)



Notes: Based on administrative student data. This graph shows the shares of students in the G8-reformed track relative to all *Gymnasium* students (with and without reform) for specific age groups. The students share is based on the G8-reformed students in each grade level, but refers to the corresponding age groups in the respective grade levels.

Figure 2.1 shows that the *G8* reform has affected almost all *Gymnasium* students, with over 80% of academic high school students were affected since 2011.

Since the crime data used in this analysis is mostly aggregated over two-year age cohorts, I calculate the share of *G8*-reformed students based on the corresponding two year grade levels. Due to this fact, it may be that only one half of a two-year age cohort was affected by the reform. Furthermore, only the 18-year-olds in the three-year age group of 18 to 20-year-olds could be affected by the reform, which reduces the maximum reform impact to one third of this three-year age cohort. To capture the maximum reform impact for further analysis, I construct a *G8* reform dummy which takes the value of 1 when the *Gymnasium* track of a two-year age group was subject to the *G8* reform in a given year and *Land*. The *G8* reform dummy takes the value of

$\frac{1}{2}$ when only one age cohort out of the two-year age group is affected in a given year.[15] The dummy variable takes the value of 0 for all other (not reformed) age groups in a given year.

2.2.4 Implications of the G8 Reform for Crime

Theoretically, it is not clear whether the *G8* reform would increase or decrease the crimes committed by affected students. Students could exhibit higher stress measures due to an increased workload, in comparison with students receiving the standard curriculum. One way to cope with the increased stress, at least in the short run, could be escapism through increased drug consumption. Violent crimes could also increase if students act out through violent behavior or are more short-tempered and engage more frequently in physical conflicts as a result of increased stress brought on by the *G8* reform. However, Milde-Busch et al. (2010) show that stress measures for reform-affected students are not significantly higher than those of high school students prior to the *G8* reform.[16]

Alternatively, increased instructional hours leave students with less residual time for committing crimes. This form of incapacitation means that students cannot commit crimes outside the school as long as they are in school. A similar effect may occur when students invest more time in studying at home to cope with the increased curriculum content, a form of self-incapacitation.[17]

While it seems theoretically feasible to just shift the crime to the after-school time of day or the weekends, the *G8* reform reduced the residual spare time of students and therefore hampered criminal behavior through simple displacement. In general, crimes and drug consumption can increase within the school environment too simply because the lengthened school day offers more opportunities to commit crimes during school hours.[18] However, this effect should be less relevant for high-performing students who have an educational aspiration and pursue an academic school track. Furthermore, teacher supervision makes committing crimes difficult in the school, including recess

[15] This can be the case in the introduction phase of the *G8* reform, when only the younger or lower age cohort of a two-year age group was affected. Furthermore, the *G8* reform dummy takes the value of $\frac{1}{3}$ to account for the reform-affected 18-year-old students in the 18-20 age cohort.

[16] In fact, Milde-Busch et al. (2010) find that the stress measures of *Gymnasium* students are high irrespective of the *G8* reform.

[17] Given the age of a student, this form of self-incapacitation may even be enforced by the parents, especially for younger students.

[18] Luallen (2006) finds that this effect is present in the U.S. for violent crime, but not for other types of crime.

and small breaks within school days. I will assess this potential effect for cannabis consumption with a student drug survey in Section 2.5.3.

2.3 Data

To study the *G8* reform's impact on crime, I compile various datasets that record crimes and substance abuse. To link the crime data with the *G8*-reformed students, I rely on yearly student enrollment data which records the number of students in different school tracks and grades. Additionally, I gather annual information from the Federal Statistical Office to use as control variables, including the population size of an age cohort, unemployment rate, youth unemployment rate, and police expenditures.

2.3.1 Police Crime Statistics (PKS)

The police crime statistics (PKS) used in this study is administered by the federal criminal police office.[19] These data allow for a comparison of crime rates among all *Länder* in Germany since 1993. An annual sample of the data covers all offenders, their criminal charges, gender, and corresponding age group. The recorded crime charges are based on police arrests rather than on criminal convictions, which might differ. I use these data until the most recent wave in 2012 and observe the groups with the following age cohorts (in years): 10-11, 12-13, 14-15, 16-17, 18-20, and 21-22. The population between the ages 10 and 18 can be, in general, subject to the *G8* reform, whereas those aged 19 to 22 years are not and serve only for further analysis.[20]

The detailed description of law violations allow me to aggregate categories of violent-, property-, and drug-related crimes. Aggregated violent crimes include assaults, homicides, and robberies. Aggregated property crimes capture any forms of theft. Aggregated drug-related crimes include all possessions or trades of illegal substances and any crimes associated with obtaining drugs. The fact that the police only reports crimes where a charge occurs, and that the true crime rate is probably higher, is not a serious measurement error as long as the fraction of reported crimes with respect

[19]Source: PKS Bundeskriminalamt, 1998-2012. Data license Germany,- attribution Version 2.0. www.bka.de/DE/Publikationen/PolizeilicheKriminalstatistik

[20]The age cohorts of the years 19-22 make it possible to analyze potential catch-up effects of former *G8* track students after their school career.

to the true crime rate is not affected by the *G8* reform. Furthermore, the police can only charge a crime in the data if the suspect is known.[21]

2.3.2 Student Enrollment Data

To link youth crime rates with student data, I gather the student enrollment share in the *G8* and *G9* regimes, based on the Annual Report of general education from the Federal Statistical Office of Germany. These data provide the number of students in each school type for each school grade in the different *Länder* beginning in school year 1998/1999 until 2012/2013.[22] These data also allow me to capture the share of students in the old *G9* regime and the new *G8* regime, which is necessary to link each to age group-specific crime rates. Since it is not possible to identify the actual age of students within one grade in a given school year, I use the legal age at school start to determine the grade level of the age groups in the youth crime data.

Due to the half-year shift in the school year with respect to the calendar year, I assign half of the students in the 5$^{\text{th}}$ grade as 10-year-olds and the other half as 11-year-olds. This results in a graduation age of 17 and 18 years, respectively, when students finish the *G8*-reformed *Gymnasium* following completion of the 12$^{\text{th}}$ grade. This does not account for grade repetitions by affected students. Huebener and Marcus (2015) find that the repetition rates in the last three grades before graduation have increased due to the *G8* reform, which could bias the results for older students. However, grade repetition is least frequent in the *Gymnasium* among the traditional secondary school branches. To merge the two year groups' crime data, I aggregate the students' school enrollment data. To account for potential correlation of residuals within *Länder* and across age cohorts, I cluster the standard errors for all models estimated with the PKS data on the *Land* level.[23]

[21] The overall crime clearance ratio is, on average, 55%, but varies strongly with specific crimes. The drug-related and violent crime clearance ratios are over 95% and 80%, respectively.

[22] Data are missing in the school year 2000/2001 for Thuringia and Saxony-Anhalt, which I interpolate with the average number of students in the previous and following school years.

[23] The results for the small number of clusters due to only 16 federal *Länder* are confirmed by regressions with a wild bootstrap procedure and other cluster units (i.e., interaction of *Land* and birth cohort).

2.3.3 Schulbus Survey Data

The Schulbus Survey, a study on substance abuse among adolescents in the *Land* of Hamburg, was conducted between 2004 and the most recent wave of 2012, and covers students between the ages of 14 and 17 years. In total, the sample covers the years 2004, 2005, 2007, 2009, and 2012. The introduction of the *G8* reform in Hamburg in the school year 2002/2003 only with the 5^{th} grade ensures that I have a reasonable number of older *Gymnasium* students that were not affected. All fourteen-year-old *Gymnasium* students up to 2005 were already too old to be affected, whereas the fourteen- and fifteen-year-old *Gymnasium* students in 2007 were the first ones affected by the *G8* reform; from 2009 on, all *Gymnasium* students younger than seventeen-years-old were subject to the *G8* reform. The survey is a repeated cross-section sample of secondary school students and explores students' general substance abuse, whether drug experiences were gained within the school environment, and the prevalence of substance abuse within peer-groups. The original sample consists of 5,508 students in the different implementation waves. I drop the students whose place of residence is unclear (405 observations) and those students who are enrolled at schools in surrounding *Länder* (169 observations).[24] The data offer a self-assessed cannabis addiction measure and questions with respect to drug prevalence in school, peer-groups, and life in general.

2.4 Identification and Estimation Strategy

To estimate the effect of an intensified curriculum in affected high schools on youth crime rates, I define the crime rate (CR) of an age group (i) in a *Land* (s) for a given year (t):

$$CR_{ist} = ln(\frac{Records_{ist}}{Population_{ist}})$$

The crime rate is the logarithm of the number of offenders divided by the corresponding population of the age group in a *Land* for a given year. Due to the fact, that the crime data is based on a two-year age group I sum up the share of affected students in groups of two grade levels respectively.

[24] In general, these students could be used as a control group also because of the later implementation in the surrounding *Länder*. However, the data do not differentiate between the surrounding *Länder* of Hamburg.

First, I will show baseline estimates with a *G8* reform dummy following the approach by Chevalier and Marie (2013). The *G8* reform dummy indicates when a *Gymnasium* track of an age cohort was subject to the *G8* reform in a given year and *Land*. In a further step, I regress the crime rates on the actual share of students within the new *G8* track to estimate the effect of intensified schooling on crime. The variable $G8\_share_{ist}$ captures the share of students in the 'new high school regime' relative to all adolescents in this age group. I rely on three different specifications to assess the causal impact of the intensified curriculum on crime rates.

The basic specification has the following structure:

$$CR_{ist} = \beta G8\_share_{ist} + \sum_i \gamma_i Age_i + \sum_t \mu_t Year_t + \sum_s \alpha_s D_s + \epsilon_{ist} \tag{1}$$

The variables Age_i account for the fixed effects of each age group, $Year_t$ absorbs all year-specific effects, and D_s captures *Länder* fixed effects.

Specification 2 adds control variables captured by the matrix X_{st}:

$$CR_{ist} = \beta G8\_share_{ist} + \sum_i \gamma_i Age_i + \sum_t \mu_t Year_t + \sum_s \alpha_s D_s + \varphi X_{st} + \epsilon_{ist} \tag{2}$$

The additional control variables account for the *Land*'s level of police expenditures, the youth unemployment rates for people under the age of 25, and the overall unemployment rates for each year in each *Land*. Specification 3 adds *Land*-specific time trends in linear and quadratic forms, $\sum\sum T_t D_s$ and $\sum\sum T_t^2 D_s$:

$$\begin{aligned} CR_{ist} = \beta G8\_share_{ist} + \sum_i \gamma_i Age_i + \sum_t \mu_t Year_t + \sum_s \alpha_s D_s + \varphi X_{st} \\ + \sum_s \sum_t \delta_s T_t D_s + \sum_s \sum_t \lambda_s T_t^2 D_s + \epsilon_{ist} \end{aligned} \tag{3}$$

The staggered implementation of the *G8* reform allows for an identification of the reform's impact on crime via (1) differences over time, (2) within *Länder*, and (3) in the age groups' proportion of students affected by the *G8* reform. I use this variation to apply a difference-in-difference approach and assess the causal impact of lengthening the school day on crime outcomes. The sample population comprises of students between the ages of 10 and 22 from all *Länder* between the school years 1998/99 and 2012/13.

All results of the estimates with the *G8* reform dummy and the actual *G8*-reformed share of students can be interpreted as semi-elasticities, based on a marginal increase of students in the reformed *G8* track: a one percentage point increase of affected students in the *G8* track triggers a β increase in the specific crime rate. The standard errors are clustered on the *Land* level.[25]

In the evaluation of the student drug survey, I estimate the effect of the *G8* reform on several drug-related binary outcomes using linear probability models. The *G8* reform dummy is equal to 1 for all *Gymnasium* students after the *G8* reform and 0 for pre-reform (*G9*) *Gymnasium* students and students from other school tracks.

2.5 Results

2.5.1 Crime

We now turn to the econometric analysis of crime data and start with treatment of the *G8* reform dummy as a baseline estimation. The underlying sample for this analysis comprises the population aged 10 to 22 years in all *Länder* (16) between 1998 and 2012. Table 2.2 includes the regression results from the estimations with a *G8* reform dummy that assumes the full age cohort is affected by the *G8* reform. Panel (A) of Table 2.2 presents the *G8* reform impact on the total crime rate. Specification (1) includes *Land*-, year-, and age group-fixed effects; Specification (2) adds police expenditure and the overall- (youth-) unemployment rate; Specification (3) adds linear and quadratic *Land* time trends. Specification (1) shows a small negative effect of -0.05 with the *G8* reform dummy, which is statistically significant at the 10% level. Specification (2) does not show a statistically significant estimate. However, in Specification (3) with the full set of control variables, one sees a point estimate of -0.06, which is statistically significant at the 5% level. This estimate suggests that the overall crime rate would decrease by 6% if the full age cohort were affected by the *G8* reform.

Panel (B) includes the regression results for drug-related crimes. All Specifications (1)-(3) show a statistically negative effect of the *G8* reform on drug-related crimes. Specification (1) estimates a decrease of -0.20 due to the *G8* reform, which is statistically significant at the 5% level. Specification (2) shows a slightly larger drop by

[25] In the case of a *Land* and cohorts interaction as the cluster unit, regression results deliver smaller standard errors than just the *Land* as a cluster unit. Therefore, I use only the *Land* level as a cluster unit to rely on the more conservative estimates. For further information, see Section 2.5.2 for a discussion of the standard errors.

Table 2.2: Reform dummy impact on different crime rates

	(1)	(2)	(3)
(A)		Overall Crime Rate	
Reform	-0.053*	-0.035	-0.063**
	(0.025)	(0.024)	(0.023)
Observations	1,440	1,344	1,344
R^2	0.272	0.222	0.349
(B)		Drug-related Crime Rate	
Reform	-0.202**	-0.241***	-0.143**
	(0.069)	(0.073)	(0.066)
Observations	1,376	1,286	1,286
R^2	0.353	0.381	0.460
(C)		Violent Crime Rate	
Reform	-0.094*	-0.043	-0.091*
	(0.044)	(0.044)	(0.046)
Observations	1,440	1,344	1,344
R^2	0.325	0.364	0.479
(D)		Property Crime Rate	
Reform	-0.009	0.006	-0.017
	(0.024)	(0.025)	(0.027)
Observations	1,440	1,344	1,344
R^2	0.685	0.681	0.725
Land, Year, and Age Group Dummies	Yes	Yes	Yes
Police Expenditure	No	Yes	Yes
(Youth) Unemployment	No	Yes	Yes
Land Specific Time Trends	No	No	Yes
Number of clusters	16	16	16

Notes: Based on PKS and administrative student data with the age groups from 10 to 22. Observation period 1998 - 2012 for specification 1 and 1998 - 2011 for specification 2 and 3. Panel B looses observation due to zero incidences of drug-related crimes in some *Länder* for a few age groups in certain years. All specifications control for the absolute population in a given age group. Standard errors are clustered on the *Land* level and reported in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

−0.24, with statistical significance at the 1% level. Specification (3) shows a slightly smaller point estimate of −0.14, but is still statistically significant at the 5% level. The effect on the violent crime rate is shown in Panel (C). Again, one sees a negative influence of the *G8* reform on the violent crime rate. Specifications (1) and (3) have virtually the same point estimate at −0.09, and are statistically significant at the 10% level. Specification (2) fails to estimate a statistically significant effect, but shows the same negative relation as Specifications (1) and (3). Panel (D) includes the *G8* reform impact on the property crime rate. The point estimates are not statistically significant and are close to zero, which suggests no effect of the *G8* reform on property crimes.

To further analyze the negative effect on drug-related and violent crimes, we turn now to the *G8* reform impact measured with the student's share in the reformed *G8* track (Share in *G8* track) on student crime rates. The independent variable in Table 2.3 captures the percentage of affected students within one age group. Panel (A) of Table 2.3 presents the effect on the overall crime rate. Specification (1) and the strictest Specification (3) estimate a decrease in the overall crime rate of −0.14 (statistically significant at the 5% and 10% levels, respectively) due to the reform affected *G8* students. One can interpret the effect as follows: for each additional percentage point of affected students, the overall crime rate declines by 0.14%. The estimate from Specification (2), with fixed effects and controls for police expenditures, unemployment rates, and youth unemployment rates, diminishes to −0.08 and loses statistical significance.

In a next step, I analyze the *G8* reform's impact on different types of crime rates. Table 2.3 reports in Panel (B) the effect of the *G8* reform on drug-related crimes. Again, Specification (1), which is controlling for *Land*, age, and year fixed effects, indicates a drop in drug-related crimes of −0.62 due to the *G8* reform. This effect is statistically significant at the 1% level. The effect in Specification (2) even increases to −0.78 (still statistically significant at the 1% level) when *Land* -specific control variables are included. Specification (3), with additional *Land*-specific time trends, again shows an estimate of −0.59 (statistically significant at the 1% level). This is a huge effect with respect to the magnitude. Given that on average, one-third of a student's age group attends a *Gymnasium*, the *G8* reform reduces the drug-related crimes by 20%. With respect to the absolute drop in the drug-related crime rate, this 20% decrease relates to a 0.11 percentage points reduction relative to the average drug-related crime rate of 0.55 percentage points.

Table 2.3: *G8* reform impact on different crime rates

	(1)	(2)	(3)
(A)	Overall Crime Rate		
Share in G8 Track	-0.135**	-0.078	-0.135*
	(0.061)	(0.066)	(0.077)
Observations	1,440	1,344	1,344
R^2	0.274	0.221	0.345
(B)	Drug-related Crime Rate		
Share in G8 Track	-0.617***	-0.778***	-0.585***
	(0.177)	(0.197)	(0.135)
Observations	1,376	1,286	1,286
R^2	0.355	0.387	0.467
(C)	Violent Crime Rate		
Share in G8 Track	-0.362***	-0.216*	-0.333***
	(0.099)	(0.102)	(0.101)
Observations	1,440	1,344	1,344
R^2	0.340	0.372	0.487
(D)	Property Crime Rate		
Share in G8 Track	0.018	0.074	0.038
	(0.063)	(0.071)	(0.086)
Observations	1,440	1,344	1,344
R^2	0.685	0.682	0.725
Land, Year, and Age Group Dummies	Yes	Yes	Yes
Police Expenditure	No	Yes	Yes
(Youth) Unemployment	No	Yes	Yes
Land Specific Time Trends	No	No	Yes
Number of clusters	16	16	16

Notes: Based on PKS and administrative student data with the age groups from 10 to 22. Observation period 1998 - 2012 for specification 1 and 1998 - 2011 for specification 2 and 3. Panel B looses observation due to zero incidences of drug-related crimes in some *Länder* for a few age groups in certain years. All specifications control for the absolute population in a given age group. Standard errors are clustered on the *Land* level and reported in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

Panel (C) of Table 2.3 shows regression results from the share of *G8* track students and violent crimes. All Specifications (1)-(3) show a negative sign in the range of -0.22 to -0.36 and are statistically significant. Specification (1) shows the highest negative impact of -0.36 and is statistically significant at the 1% level. Specification (2), with further control variables, estimates a slightly lower effect of -0.22, which is only statistically significant at the 10% level. Specification (3), with the full set of control variables and time trends, shows a higher point estimate of -0.33 and is again statistically significant at the 1% level. The effect on violent crime in this estimation confirms the negative relationship, which was estimated before with the simple treatment dummy. Panel (D) of Table 2.3 reports the estimated effect of the *G8* reform on property crimes. The point estimates over all Specifications are positive; however, none of the estimates are statistically significant due to large standard errors.

Table 2.4: Age group specific *G8* reform impact on drug-related crimes

	(1)	(2)	(3)
	Drug-related Crime Rate		
Share in G8 at Age 10-11	-2.343**	-1.903*	-1.283
	(0.969)	(1.046)	(1.001)
Share in G8 at Age 12-13	-0.907***	-1.032***	-0.807***
	(0.286)	(0.298)	(0.270)
Share in G8 at Age 14-15	-0.406**	-0.592***	-0.460***
	(0.167)	(0.191)	(0.139)
Share in G8 at Age 16-17	-0.319	-0.469*	-0.365*
	(0.206)	(0.227)	(0.190)
Share in G8 at Age 18-20	-0.252	-0.372	-0.281
	(0.719)	(0.922)	(0.791)
Land, Year, and Age Group Dummies	Yes	Yes	Yes
Police Expenditure	No	Yes	Yes
(Youth) Unemployment	No	Yes	Yes
Land Specific Time Trends	No	No	Yes
Observations	1,376	1,286	1,286
R^2	0.3751	0.3969	0.4720
Number of clusters	16	16	16

Notes: Based on PKS and administrative student data with the age groups from 10 to 22. Observation period 1998 - 2012 for specification 1 and 1998 - 2011 for specification 2 and 3. All specifications control for the absolute population in a given age group. Standard errors are clustered on the *Land* level and reported in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

An analysis of different effects of the affected age groups with respect to drug-related crimes is shown in Table 2.4. The table includes effects from one single regression with a set of variables capturing the share of students in the *G8* track of each age group. The biggest effect is present for the youngest age group (10-11) in Specification (1) with a point estimate of -2.34, which is statistically significant at the 5% level. However, the effect diminishes slightly in Specifications (2) and (3) and is no longer statistically significant in the strictest Specification (3). Robust estimates with respect to the different specifications are present for the age groups (12-13) and (14-15), which indicates an almost twice as high effect for the younger group (12-13). The effect of this group (12-13) varies between -0.81 and -1.03 and is statistically significant at the 1% level. The slightly older age group (14-15) shows a drop from -0.41 to -0.59, with higher negative effects in Specifications (2) and (3). Whereas the effect in Specification (1) is statistically significant at the 5% level, Specifications (2) and (3) show a statistical significance of the effect at the 1% level. The point estimates of the older age group (16-17) are negative too; however, only Specifications (2) and (3) yield a statistically significant effect at the 10% level for the age group (16-17). Furthermore, the effect of the share of *G8*-reformed students is not statistically significant for the oldest age group (18-20). One has to bear in mind that only the 18-year-old students in this age group were subject to the *G8* reform and only for a half-calendar year.[26] As described before, an increase in grade repetition could bias the results of these age groups (16-17 and 18-20) and one should treat these results with caution.

Table 2.5 focuses solely on the *G8* reform's impact on cannabis, which is the most prominent illegal drug in Germany and accounts for more than half of all drug-related crimes. One can expect that the *G8* reform has diverse results when discriminating between the serious crime of dealing and the rather delinquent behavior of the pure consumption of cannabis. Therefore, Table 2.5 separates between the *G8* reform impact on the rate of dealing with cannabis and the rate of consuming cannabis. Panel (A) of Table 2.5 shows the effect from the relative students intake in the *G8* track effect on the cannabis dealing rate. The *G8* reform seemed not to have an effect on the actual rate of dealers. Panel (B) of Table 2.5 presents the effect on the cannabis use rate. The negative point estimates of Specifications (1)-(3) range from -0.60 to -0.81 and are all statistically significant at the 1% level, suggesting a strong negative effect of the *G8* reform on the cannabis consumption of a student's age group.

[26]Since the German school year ends usually in June or July, the overlap of the student enrollment data and yearly crime data (based on the calendar year) is reduced to one-half.

Table 2.5: *G8* reform impact on cannabis dealing and consuming rate

	(1)	(2)	(3)
(A)	Cannabis Dealing Crime Rate		
Share in G8 Track	-0.067	-0.248	-0.238
	(0.260)	(0.294)	(0.268)
Observations	1,218	1,138	1,138
R^2	0.323	0.337	0.475
(B)	Cannabis Consuming Crime Rate		
Share in G8 Track	-0.602***	-0.812***	-0.719***
	(0.172)	(0.199)	(0.172)
Observations	1,350	1,263	1,263
R^2	0.340	0.367	0.442
Land, Year, and Age Group Dummies	Yes	Yes	Yes
Police Expenditure	No	Yes	Yes
(Youth) Unemployment	No	Yes	Yes
Land Specific Time Trends	No	No	Yes
Number of clusters	16	16	16

Notes: Based on PKS and administrative student data with the age groups from 10 to 22. Observation period 1998 - 2012 for specification 1 and 1998 - 2011 for specification 2 and 3. Panel A looses observation due to zero incidences of crimes related to dealing with cannabis in some *Länder* for a few age groups in certain years. All specifications control for the absolute population in a given age group. Standard errors are clustered on the *Land* level and reported in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

2.5.2 Sensitivity Analysis

The fact that the share of other school types has varied over recent years, as the share of students in *Gymnasium* has increased, makes it possible that newer trends in the school type composition are correlated with the *G8* reform. Therefore, I estimate the effects from *G8*-reformed students on drug-related crimes while controlling for other shares of school types. The share of other school types (Hauptschule, Realschule, and comprehensive schools) or the share of the non-reformed *G9* track do not drive the estimates of the *G8*-reformed students. Estimations with the share of *G8* and *G9* track students occasionally deliver a higher level of statistical significance for the effect of the share in the *G8* track; however, the point estimates of the students in the *G9* track are closer to zero and never statistically significant.

All estimation results from the share of students in the *G8* track and the police crime data show the same level of significance when the standard errors are clustered on the interaction between *Land* and cohorts. Due to lower standard errors of this and other cluster units, the presented results with the *Land* as cluster delivers the most conservative results (largest standard errors) as suggested by Bertrand et al.(2004). The fact, that the *Land* and cohort cluster units deliver larger standard errors suggests that serial correlation seems to be not present in the underlying panel data.

However, the problem of too few clusters due to only sixteen German Länder might generate over-rejection resulting in too narrow confidence intervals as discussed in Cameron and Miller (2015). Therefore, I perform all regression results as a robustness check with a wild bootstrap procedure. The estimations confirm the significance of the results derived from regressions with the police crime statistics and student drug survey and do not deliver wider/broader confidence intervals. Regression results with only the *G8* reform-affected sample [without the unaffected age group (21-22)], do not change the estimates of the *G8* reform effect on the crime behavior of younger *G8*-affected students.

To estimate if a 'catch-up' effect takes place in the years after the graduation of *G8* reform-affected students, I estimate a model where another treatment dummy indicates former *G8* track students. This treatment dummy follows students who attended the *G8*-reformed *Gymnasium* after their school career is finished and they potentially pursue a tertiary education or an apprenticeship. This after school reform dummy follows *G8* reform-affected age cohorts after graduation from the *Gymnasium*, when they are between 19- and 22-years-old. One could expect to see an effect if former *G8* track students use the time after graduation to engage in criminal activity, because the time and commitment constraints of *G8*-reformed *Gymnasium* prevented them from doing so earlier. However, given that not all *Länder* finished so far the *G8* reform implementation and others just released their first double graduation cohorts, these results should be treated with caution due to a low number of former *G8*-affected students. With respect to drug-related crimes, it appears that no catch-up effect is taking place. Occasionally, the negative impact of former *G8* track students on drug-related crimes seems to be prolonged after high school graduation; however, the effect is not robust over all specifications. The negative effect of the *G8* reform on violent crimes could face a catch-up effect after high school graduation. The impact of former

G8 track students with the after school reform dummy on violent crimes is positive and comparable with the negative violent crime effect of the share of *G8* track students with respect to the magnitude.[27]

To deal with the potential problem of grade repetition as discussed by Huebener and Marcus (2015), I estimate all regression results from the crime data with an additional control for the share of grade repetition, which was affected at least partially by the reform. I use one variable for the grade repetition of *Gymnasium* students and another variable for all other school types. These variables control for the lagged (previous year) grade repetition rate of the school types in the next higher grade level. These control variables take into account that grade repeating students show up in the crime data in the next higher age cohort even when they remain in the grade level from the previous year. These regression results suggest that the presented *G8*-reform effect on crime is not driven by the increased grade repetition of reform-affected students.[28]

2.5.3 Survey Data

We turn now from the effects of the official crime statistics to the impact of the *G8* reform on the student drug survey to provide further evidence of reduced drug-consumption. To make sure that the effect was mainly driven by the affected students in the high school track and not by changes in the drug consumption of students in other school tracks, I now turn to a student drug survey. Potentially, one could think that the effect from the *G8*-reformed students exerts an additional effect on the peer group of *G8*-students if the peer group is distributed across different school types.

The sample consists of a repeated cross-section and covers students from reformed as well as non-reformed schools of different types. The *G8* reform took place within the third wave of the survey and affected all *Gymnasium* students in the last two waves. The Schulbus survey data from the *Land* of Hamburg is evaluated using a linear probability model, which assigns all *Gymnasium* students after the *G8* reform with a reform dummy.

Table 2.6 shows the impact of the *G8* reform on cannabis consumption within the school environment. The definition of school environment includes the school grounds and external school trips. The first three columns capture the estimates of the full-sample and columns four to six restrict the sample to students younger than sixteen

[27]These regression results are reported in Table 2.10 in the Appendix.
[28]Results are available from the author upon request.

years. This assures that the sample is not driven by sample selection of school dropouts and early starters in the job market. Furthermore, this age group faced the biggest increase of 'shifted' school hours. The negative relation of the *G8* reform on the consumption of cannabis within the school environment is statistically significant in the specification without controls [(1) and (4)] and some control variables [(2) and (5)] for the full and age-restricted samples, respectively. Although the relation is not significant, it is negative in the specifications (3) and (6) which include the full set of control variables.

If this effect were positive, then the drop in the police crime statistics could be explained by a shift of crime to the school environment due to the longer school hours. This could be the case if delinquencies within the school environment are more likely to be sanctioned by teachers rather than the police to reduce administrative duties or to preserve the school's reputation. Furthermore, police patrols, which are less prevalent in schoolyards than outside the school environment, could result in fewer crimes being detected by the police. The evidence clearly rejects that the consumption of cannabis was shifted into the school environment. In fact the (insignificant) negative point estimates indicate that cannabis consumption within the school has slightly decreased as instructional hours increased.

Table 2.6: *G8* reform impact on cannabis using in school within the last year

	(1)	(2)	(3)	(4)	(5)	(6)
	Cannabis usage in school environment					
Reform	-0.040***	-0.071***	-0.029	-0.049***	-0.059***	-0.029
	(0.011)	(0.017)	(0.019)	(0.010)	(0.019)	(0.025)
Male	0.053***	0.058***	0.058***	0.034***	0.035***	0.036***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Fixed Effects for ...						
Age Groups	No	Yes	Yes	No	Yes	Yes
School Types	No	Yes	Yes	No	Yes	Yes
Waves	No	No	Yes	No	No	Yes
Districts	No	No	Yes	No	No	Yes
Sample	$age: 14-17$			$age: 14-15$		
Observations	3,005	2,989	2,989	1,741	1,733	1,733
R^2	0.004	0.042	0.057	0.008	0.024	0.044

Notes: Based on weighted survey data from Schulbus waves. Observation period 2004 - 2012. Cannabis in school is a dummy variable equal to one if at least once cannabis was consumed in the school environment within the last 12 months, zero otherwise. The definition of school environment includes, among others, (breaks at) the schoolyard and school excursions. Robust standard errors are reported in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

The negative effect of the *G8* reform on drug consumption that I identified in the crime statistics is also present in the student drug survey. Table 2.7 shows the regression results for an indicator of cannabis addiction as a dependent variable.[29]

Table 2.7: *G8* reform impact on cannabis addiction

	(1)	(2)	(3)	(4)	(5)	(6)
	Cannabis addiction					
Reform	-0.024***	0.032	0.008	-0.032***	-0.007	-0.024*
	(0.009)	(0.019)	(0.022)	(0.010)	(0.013)	(0.014)
Male	0.042***	0.042***	0.043***	0.040***	0.039***	0.041***
	(0.009)	(0.009)	(0.009)	(0.011)	(0.011)	(0.011)
Fixed Effects for ...						
Age Groups	No	Yes	Yes	No	Yes	Yes
School Types	No	Yes	Yes	No	Yes	Yes
Waves	No	No	Yes	No	No	Yes
Districts	No	No	Yes	No	No	Yes
Sample	$age : 14-17$			$age : 14-15$		
Observations	2,345	2,329	2,329	1,408	1,400	1,400
R^2	0.003	0.023	0.027	0.007	0.021	0.030

Notes: Based on weighted survey data from Schulbus waves. Observation period 2004 - 2012. Cannabis addiction is observed if at least 2 items of the Severity of Dependence Scale apply. Robust standard errors are reported in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

For the full sample, only the first specification (with only a male dummy indicator) shows a statistically negative relationship. For the sample restriction to the age group (14-15), Specifications (4)-(6) indicate a negative sign that ranges from -0.01 to -0.03. Specification (6), with the full set of control variables, suggests that the rate of cannabis-addicted students decreased by -0.02% for 14- to 15-year-old students after introduction of the *G8* reform.[30]

Table 2.8 shows the effect on the cannabis prevalence within the peer groups of *G8*-reformed students. The outcome variable is a dummy equal to one whenever the student indicates that half or more of their peer group has experiences with cannabis. Again, one finds that the *G8* reform goes in line with a decreasing prevalence of

[29]Cannabis addiction is measured with a binary variable that is equal to one if at least two out of five items of cannabis addiction symptoms apply. The 5 symptoms are measured on the Severity of Dependence Scale and defined in the Appendix.

[30]The results are robust when only the sub-sample of males is considered for the regression.

cannabis within the peer group. Although the Specifications (1) and (4) as well as (2) and (5) show a highly statistically significant relation, significance vanishes in the most strict Specifications (3) and (6).

Table 2.8: *G8* reform impact on cannabis prevalence in peer-group

	(1)	(2)	(3)	(4)	(5)	(6)
	Cannabis prevalence in peer-group					
Reform	-0.088***	-0.061***	-0.014	-0.126***	-0.079***	-0.032
	(0.015)	(0.020)	(0.029)	(0.016)	(0.027)	(0.040)
Male	0.047***	0.047***	0.046***	0.042**	0.039**	0.036**
	(0.015)	(0.014)	(0.014)	(0.017)	(0.017)	(0.017)
Fixed Effects for ...						
Age Groups	No	Yes	Yes	No	Yes	Yes
School Types	No	Yes	Yes	No	Yes	Yes
Waves	No	No	Yes	No	No	Yes
Districts	No	No	Yes	No	No	Yes
Sample	$age: 14-17$			$age: 14-15$		
Observations	3,652	3,636	3,636	2,094	2,086	2,086
R^2	0.009	0.056	0.076	0.021	0.056	0.066

Notes: Based on weighted survey data from Schulbus waves. Observation period 2004 - 2012. Cannabis prevalence in peer-group is a dummy variable equal to one if at least half of the peer-group is consuming cannabis, zero otherwise. Robust standard errors are reported in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

Table 2.9 shows the *G8* reform impact on the monthly prevalence of cannabis consumption in the top panel and the lifetime prevalence in the bottom panel. The negative and statistically significant pattern of estimates from the less strict specifications is present here, too. The fact that the point estimates for the lifetime prevalence is more than twice as high could explain that the *G8* reform impact is more likely to have an effect on the extensive margin and less on the intensive margin. In other words, the *G8* reform seemed to have stopped students from starting to use cannabis or delayed the starting age rather than making them just consume less.

Table 2.9: *G8* reform impact on cannabis prevalence

	(1)	(2)	(3)	(4)	(5)	(6)
	Cannabis 30-day prevalence					
Reform	-0.020*	-0.005	0.017	-0.038***	-0.017	0.003
	(0.012)	(0.016)	(0.023)	(0.013)	(0.021)	(0.030)
Male	0.074***	0.075***	0.074***	0.067***	0.065***	0.064***
	(0.011)	(0.011)	(0.011)	(0.013)	(0.013)	(0.013)
Observations	3,847	3,831	3,831	2,219	2,211	2,211
R^2	0.001	0.033	0.040	0.004	0.028	0.036
	Cannabis lifetime prevalence					
Reform	-0.066***	-0.066***	0.047	-0.104***	-0.105***	0.035
	(0.016)	(0.022)	(0.033)	(0.017)	(0.030)	(0.043)
Male	0.104***	0.109***	0.108***	0.079***	0.080***	0.080***
	(0.015)	(0.015)	(0.015)	(0.018)	(0.018)	(0.017)
Observations	3,849	3,833	3,833	2,221	2,213	2,213
R^2	0.005	0.062	0.081	0.013	0.041	0.059
Fixed Effects for ...						
Age Groups	No	Yes	Yes	No	Yes	Yes
School Types	No	Yes	Yes	No	Yes	Yes
Waves	No	No	Yes	No	No	Yes
Districts	No	No	Yes	No	No	Yes
Sample	$age: 14-17$			$age: 14-15$		

Notes: Based on Schulbus. Observation period 2004 - 2012. Robust standard errors are reported in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

2.6 Conclusion

This analysis leads to the conclusion that the moderate decline in the overall crime rate is due to the stronger decline in drug-related and violent crimes. However, this reduction can be mainly explained by a drop in delinquencies of the cannabis-user rate rather than the use of hard drugs or even drug dealers. Furthermore, the analysis provides evidence that lengthening the school day at lower grades, when students are aged 12 to 15 years, reduces drug-related crimes (at least for the high school track). The drop in drug-related crimes is mainly attributable to cannabis possession and no effect on the cannabis dealing rate is present. Further survey evidence clearly links the decrease in cannabis consumption to *G8*-affected students and rejects the hypothesis of drug consumption shifted into school hours. In fact, the survey evidence even suggests that cannabis consumption within the school decreases due to increased instructional hours. Furthermore, the prevalence of cannabis within the peer group decreases after the reform. The analysis of the student survey results indicate that the *G8* reform has stopped students from using cannabis or at least has delayed the starting age rather than just reducing the consumption.

A direct analysis of the eliminated school year (old grade 13) on crime is not feasible because this age (grade) level cannot be carefully identified within the underlying crime data. In general, I cannot rule out that the decreased crime measures during school time rise again during the eliminated school year, which is the first year after graduation. However, the first cohorts of graduates that went through the new *G8* reform system do not show a tendency to catch-up with their drug consumption in the first years after their school career. A catch-up effect of violent crime seems to be present for these cohorts. Further analysis is necessary to identify the direct effect of the eliminated school year.

Appendix

Tables

Table 2.10: *G8* reform impact on former students' crime rate

	(1)	(2)	(3)
(A)	Drug-related Crime Rate		
G8 dummy	-0.267***	-0.267***	-0.012
after school	(0.079)	(0.088)	(0.082)
Observations	1,376	1,286	1,286
R^2	0.342	0.365	0.456
(B)	Violent Crime Rate		
G8 dummy	0.131***	0.137***	0.086**
after school	(0.021)	(0.023)	(0.039)
Observations	1,440	1,344	1,344
R^2	0.316	0.368	0.469
Land, Year, and Age Group Dummies	Yes	Yes	Yes
Police Expenditure	No	Yes	Yes
(Youth) Unemployment	No	Yes	Yes
Land Specific Time Trends	No	No	Yes
Number of clusters	16	16	16

Notes: Based on PKS and administrative student data with the age groups from 10 to 22. Observation period 1998 - 2012 for specification 1 and 1998 - 2011 for specification 2 and 3. Panel A looses observation due to zero incidences of crimes related to dealing with cannabis in some *Länder* for a few age groups in certain years. All specifications control for the absolute population in a given age group. Standard errors are clustered on the *Land* level and reported in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

Definition: Severity of Dependence Scale SDS

1 Did you ever think your use of cannabis was out of control?

2 Did the prospect of missing cannabis make you very anxious or worried?

3 Did you worry about your use of cannabis?

4 Did you wish you could stop your use of cannabis?

5 How difficult would you find it to stop or go without cannabis?

- Responses:
  - Item 1-4: never or almost never (0); sometimes(1); often (2); always or nearly always (3)
  - Item 5: not difficult (0); quite difficult (1), very difficult (2); impossible (3)
- The code from the responses are added and account for a cannabis addiction if the value is at least 2 according to the definition of the SDS.

Chapter 3

Bunching on the Autobahn: Speeding responses to a 'notched' penalty scheme

Co-authored with Christian Traxler & Ansgar Wohlschlegel

3.1 Introduction

Traditionally, microeconomics focuses on analyzing smooth incentive schemes. In reality, however, agents often face regulations that imply non-linear or non-convex budget sets, i.e., policies with 'kinks' or 'notches' (Kleven, 2016). While a quickly growing body of research explores such kinks and notches in taxation,[1] discontinuous policy schemes are rarely studied beyond public finance. One important domain where notches are ubiquitous is law enforcement. In many cases, fines and other penalties change discontinuously with the 'nuances' of violation of law. Numerous laws specify, for instance, that 'minor' fraud, theft, or tax evasion is punished differently than 'major' cases. The differences between minor and major cases are often defined along given monetary cutoffs (e.g., related to the damage; see Rasmusen, 1995). In a similar vein, sentencing guidelines often include discontinuous jumps at thresholds regarding the 'offense score' or the 'offender score'. Driving under the influence triggers a penalty that discontinuously increases at certain cutoff levels of blood alcohol content (Hansen, 2015).

[1] See, among many others, Saez (2010), Chetty et al. (2011), and Kleven and Waseem (2013).

Similarly, this also applies for speeding, where penalties typically increase stepwise in the speed level. The latter case forms the institutional context for our paper.

We analyze drivers' responses to a penalty scheme in which penalties increase discontinuously at certain levels of speed above the respective limit. The core of our study exploits detailed data on more than 150,000 speeding tickets recorded on the German Autobahn. Like in many other countries, fines and other penalties jump discontinuously at several speed levels (e.g., 20, 25 or 30 km/h above the speed limit). In Slemrod's (2013) terminology, drivers thus face numerous 'notches' in the penalty structure. To set the stage for our empirical analysis, we introduce a simple analytical framework in the spirit of Kleven and Waseem (2013) and study the role of notches on drivers' optimal speed choices. The analysis shows under which conditions drivers with heterogeneous tastes for speeding bunch at speed levels with notches.

Before we study the patterns of speed among the speeding tickets, we present evidence from a survey which assesses whether German drivers understand the penalty scheme's structure. The survey reveals that the majority of respondents are very knowledgeable about the scheme's stepwise shape, its discontinuous jumps and the location of the cutoffs. This finding is by no means trivial and – potentially due to the simple structure of the penalty scheme – sets our results apart from studies documenting limited knowledge (e.g., Chetty et al. 2013) and misperceptions of non-linear or non-convex budget sets (e.g., De Bartolome, 1995; Liebman and Zeckhauser, 2004; Feldman et al., 2015).[2]

Consistent with the survey evidence and the predictions from our model, we find evidence on bunching in the distribution of speeding tickets. A disproportionately high fraction of drivers are speeding exactly at or slightly below several cutoffs of the penalty scheme. Like in many tax applications (e.g., Bastani and Selin, 2014), however, bunching varies considerably along the speed distribution. For very high levels of speed (150km/h and higher), we do not detect any bunching. The observation is consistent with drivers underestimating the detection risk and the fact that driving at very high speed depletes cognitive capacities to optimally trade-off risks (Jäncke et al., 2008).

In a further step, we analyze a reform of the penalty system which increased the size of several notches (without changing the cutoffs). Our data indicate that speeders rationally responded to the reform by avoiding speed ranges which triggered significantly higher penalties after the reform. Overall, the data suggest that the reform produced a sizable shift in the speeding distribution, with a 25% drop in the fraction of cars driv-

[2] Further evidence is discussed in Chapter 2 in Congdon et al. (2011).

ing more than 20km/h above the limit. Consistent with this observation, aggregate statistics show a substantial drop in accidents and fatalities on German highways.

Our study relates to several strands of research. First, we contribute to the law and economics of speeding and speed control.[3] Given that each year around 1,000,000 lives are lost worldwide due to motor vehicle accidents (Peden et al., 2004) – with speeding being a major contributor to the number of traffic fatalities – improving our understanding of speed control policies is important. Several quasi-experimental studies document the impact of police enforcement (DeAngelo and Hansen, 2014), speed limits (van Benthem, 2015) and speeding tickets (Dusek and Traxler, 2016) on travel speed, accidents, fatalities and air pollution externalities. The present study differs from these contributions since it analyzes drivers' responses to the specific *structure* of speeding penalties.

The results from our positive analysis also carry implications for the normative debate on optimal speed limit enforcement and optimal penalties in general. At first, one might argue that a notched penalty scheme entails welfare losses: given that the externalities from speeding (accident risk, air and noise pollution, etc.) are continuously increasing with the speed level, a stepwise penalty scheme does not correspond to an efficient Pigouvian correction mechanism.[4] In the context of boundedly rational or imperfectly informed agents, however, the simplicity of the stepwise scheme might increase awareness and contributes to the good knowledge of the penalty system – a fact which is consistent with our survey evidence. Overall, the notched system could therefore be superior to a more complex penalty function that is closer to a 'true' Pigouvian scheme, but poorly understood by drivers. A further point that speaks in favor of the notched system directly relates to our bunching evidence: the fact that many drivers speed at similar speed levels right below the cutoffs tends to reduce the variance in speed. As a lower variance contributes to a reduction in the accident risk (see, e.g., Lave, 1985), bunching might be good news in itself.

Finally, in terms of methods we are the first to employ the tools from the public finance literature (e.g., Chetty et al. 2011; Kleven and Waseem 2013) to analyze responses to notched penalty schemes in law enforcement.[5] Our work therefore highlights a new field of applications for the bunching framework (Kleven, 2016). At the

[3]For early, theoretical contributions in this field see, e.g., Jondrow et al. (1983), Lavy (1985), and Graves et al. (1993).

[4]For a related discussion, see Sallee and Slemrod (2012) and, for more formal treatments, Blinder and Rosen (1985) and Gillitzer et al. (2016).

[5]A paper that is closely related to ours is the work by Sallee and Slemrod (2012), who study automakers' responses to notches in the taxation of cars.

same time, our analysis clarifies several key differences between the law enforcement and the taxation context. Most importantly, in our case, bunching is proportional to *expected* notches – the discontinuous increase in penalties weighted with the detection probability. Hence, there are two policy parameters which jointly determine the incentive to choose a corner solution: the jump in penalties at a given speed (analogous to, e.g., increases in average tax rates at certain income levels) and the risk of punishment. This latter dimension, which is not present in taxation studies but crucial if one explores notches in law enforcement, impedes the translation of bunching mass into behavioral response elasticities. The reason is that objective variation in law enforcement and subjective priors about the detection risk (speed controls) essentially add an additional layer of heterogeneity. Without common knowledge about the enforcement risk (or subjective risk assessment data), notches in penalty schemes cannot readily be used to identify behavioral elasticities.[6]

The rest of the paper is structured as follows. Section 3.2 describes the institutional framework for speeding in Germany. Section 3.3 presents evidence from our survey on drivers' knowledge of the penalty scheme's structure. Section 3.4 introduces a simple model of speeding and discusses several predictions. After presenting the data (Section 3.5), we turn to the empirical analysis of the speeding tickets in Section 3.6. The normative implications of our findings are discussed in the concluding Section 3.7.

3.2 Institutional Background

Despite common prejudices about German highways being the great dream of speeders, there are speed limits on more than 85% of the 13,000 kilometers of Autobahn. Speed limits are primarily imposed for safety reasons: high speed is the leading cause of roughly 4,000 annual traffic deaths and 400,000 annual traffic injuries in Germany. On the Autobahn, speeding is the chief cause for every second fatality.[7]

The enforcement of speed limits is based on permanently installed and on mobile speed cameras which measure the speed of passing vehicles.[8] For a speed above a certain

[6] A further, more technical difference to the taxation literature is related to the close proximity of potential bunching points. In our context, it is reasonable to consider drivers who are indifferent about speeding 20 or 25km/h above the limit. Our analysis therefore considers the joint influence of multiple, potentially inter-related notches on behavior – a point which advances and generalizes the theoretical bunching literature.

[7] Source: *Deutsches Statistisches Bundesamt.*

[8] Mobile cameras are set-up by an officer for a few hours. In addition, on-board speed measurement is carried out in unmarked cars.

level (see Section 3.5), a picture is automatically taken and the speed is recorded. Car owners are identified from the number plates and receive a ticket by mail. Penalties for speeding offenses – i.e., fines, 'penalty points' and possible driving bans – are a function of the measured speed: the speed camera's measure s is first rounded down to the next integer; a tolerance value of 3% is subtracted and the result is again rounded to the next lower integer.[9] The outcome from this so-called 'tolerance rule' (which serves as a concession to prevent appeals against speeding tickets), the speed level x, determines the penalty according to Table 3.1.

Table 3.1: Penalties for speeding at German Autobahn

Speed bracket *(speed above limit in km/h)*	Cutoff	Fines *(in euro)* pre-reform		post-reform		change	Penalty Points	Driving Ban
	x_i	f_i^{pre}	Δ_i^{pre}	f_i^{post}	Δ_i^{post}	$\frac{\Delta^{\text{post}}-\Delta^{\text{pre}}}{\Delta^{\text{pre}}}$		*(Months)*
$x \leq 10$	10	10.0	10.0	10.0	10.0	0	0	–
$10 < x \leq 15$	15	20.0	10.0	20.0	10.0	0	0	–
$15 < x \leq 20$	20	30.0	33.5	30.0	63.5	0.90	0	–
$20 < x \leq 25$	25	63.5	10.0	93.5	10.0	0	1	–
$25 < x \leq 30$	30	73.5	15.0	103.5	40.0	1.67	3	(1)
$30 < x \leq 40$	40	98.5	25.0	143.5	40.0	0.60	3	(1)
$40 < x \leq 50$	50	123.5	50.0	183.5	80.0	0.60	3	1
$50 < x \leq 60$	60	173.5	125.0	263.5	200.0	0.60	4	1
$60 < x \leq 70$	70	298.5	100.0	463.5	160.0	0.60	4	2
$70 < x$	–	398.5	–	623.5	–	–	4	3

Notes: The table presents the fines (f_i) and other penalties (penalty points and temporary driving bans) for different speed levels over the speed limit on the German Autobahn. The columns labeled by Δ^{pre} and Δ^{post} capture the pre-/post-reform notches in the fines, respectively (i.e., the increase in the fine associated with moving from a 'lower' to a 'higher' speed bracket). The duration of temporary driving bans indicated brackets are only imposed the second time a driver is detected speeding by more than 25 km/h within one year.

Monetary fines range from 10 to 623.50 euro. The fines discontinuously increase at cutoffs of x being, e.g., 20, 25, 30 or 40 km/h above the speed limit. Within each speed bracket, i.e., between two cutoffs, the fine is constant. The same holds for temporary driving bans: a one-, two- or three-month ban is imposed for speeding in the range 40–60, 60–70, and more than 70 km/h above the limit, respectively. The penalty

[9] Consider the following example: a speed camera measures $s = 140.6$km/h. The recorded speed is first rounded down to 140km/h. Thereafter, it is reduced by 3% to 135.8km/h and further rounded down to $x = 135$km/h.

point scheme follows a stepwise pattern, too. Speeding in the range 20–25, 25–50, or above 50 km/h results in one, two or three penalty points, respectively. The repeated accumulation of points, which are recorded in a register of traffic offenders, can result in the revocation of a driver's license.[10]

In 2009 there was a significant reform of the penalty schedule. Starting with February 2009, fines for speeding with more than 20 km/h above the limit were increased considerably. All other penalties (points and driving bans) remained unchanged. An overview of the fines before and after the reform as well as the stable components of the penalty system is provided in Table 3.1.

The table illustrates the key property of the penalty system: the penalty scheme is characterized by what the Public Finance literature calls 'notches' (Slemrod, 2013): discontinuous increases in fines, penalty points and/or driving bans at each speed bracket's cutoff.[11] Before 2009, speeding with e.g. 20km/h above the limit triggered a fine of 30 euro and no penalty point; for 21km/h, it was 63.50 euro and one penalty point. The table further shows that the reform not only increased the level of the fines but also the magnitude of several notches. At the 20km/h cutoff, for instance, the increase in the fine amounted to 33.50 euro before (Δ^{pre}) but 63.50 euro after the reform (Δ^{post}). The notch (in the fine level) thus increased by 90 percent. Similar increases occurred at other cutoffs.

3.3 Survey Evidence

Let us first study whether the simplicity of the stepwise penalty structure is reflected in a good knowledge of the penalty scheme. To approach this question, we conducted an online survey. The survey was implemented in June 2013 in cooperation with a professional survey company which maintains a large sample of German individuals that is representative in several dimensions (age, gender, education and occupational structure). We invited a random subset of this sample (conditional on having a driver's license) to participate in the survey. Summary statistics for the approximately 1,000

[10]Offenders who have accumulated between 14 and 17 points are obliged to participate in a costly seminar on traffic safety. Drivers that end up at 18 or more points get their driver's license revoked. Older points are deleted two years after collecting them if no additional tickets were issued since then. For a theoretical analysis of combining monetary fines with penalty points, see Bourgeon and Picard (2007).

[11]Notched penalty structures can be found in many other countries. For evidence from Italy, Spain, and the Czech Republic, see De Paola et al. (2013), Castillo-Manzano et al. (2010) and Montag (2014), respectively.

participants are provided in Table 3.2. 48% of the respondents were male and the average age was 43 years. 54% drive a car every day and more than 60% drive on the Autobahn several times a month. More than a third of the respondents have experience with the penalty system: 28% report that they were caught speeding during the last two years and 12% indicate that they hold a positive penalty point record in the register of traffic offenders (see above).

Table 3.2: Summary statistics – survey data

Variable	*Mean*	*Std. Dev.*
Age (in years)	43.324	13.986
Male	0.477	0.501
Drive car every day	0.541	0.499
Drive on Autobahn regularly	0.616	0.487
Speeding ticket within last 2 years*	0.275	0.447
Penalty point record*	0.122	0.327
Aware of tolerance rule	0.933	0.251
Speed at slightly below threshold*	0.848	0.359
Survey duration (minutes)	5.593	3.362

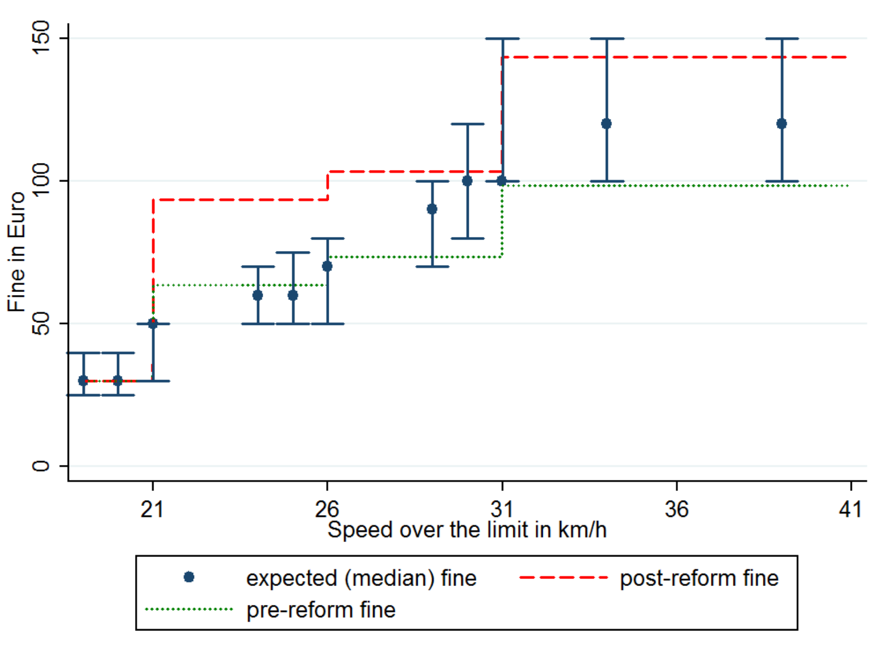
Notes: The table presents summary statistics for the online survey. The number of observations is $N = 980$. * indicates variables which are only available for sub-samples.

To elicit people's knowledge about the penalty system, we first asked survey participants to indicate the level of speeding fines for a randomly drawn sequence of speed levels (see the Supplementary Appendix for further details). We thereby obtained information on the expected fines as a function of the speed *without* mentioning the stepwise fine structure in the question. Figure 3.1 illustrates the results. The blue dots indicate the median expected fines (together with the 33rd- and 66th-percentile) for the surveyed levels of speeding. The dashed red (dotted green) line shows the actual fine for the post- (pre-) reform period. The figure provides two insights. First, respondents reveal a good knowledge of the stepwise structure and the increase of fines at the cutoffs.[12] While there is no jump in the median response at the 30/31 threshold, the lower and top terciles (as well as the average, not depicted) strongly increase at the cutoff. Second, the expected *level* of the speeding fines is much closer to the pre-reform level.

[12]The average survey duration was 6 minutes, suggesting that responses do not stem from ad-hoc online research on the questions.

A possible interpretation of this latter finding is that drivers' expectations converge only slowly to the post-reform levels.

Figure 3.1: Expected and actual fines (in Euro) for a given speed above the limit.



Note: The figure illustrates survey responses regarding the expected fines (in euro, vertical axis) for a given speed above the limit (in km/h, horizontal axis). The blue dots capture median expectation, the upper and lower 'bounds' on the blue dots indicate the 33rd- and 66th-percentile, respectively. The dashed red lines and the green dotted line depict the fines for the post- and pre-reform period, respectively.

The good understanding of the stepwise shape of the penalty function is further captured in the responses to subsequent questions which explicitly asked whether there is an increase in penalties at certain speed levels. For each of the five surveyed thresholds, the mode of the response distribution (typically accounting for half of all answers) overlaps with the true cutoff (see the Supplementary Appendix). This corroborates the first finding from above. We also asked whether drivers know about the 'tolerance rule' for computing the speed level which is relevant for determining the penalty (compare Section 3.2). 93% answered that they were aware of this rule. Among them, 36% – again the mode of the response distribution – indicated the correct rule. Hence, there is quite some variation in the expectations regarding the tolerance rule, but one out of three drivers seems to know the rule.

Overall, the survey evidence indicates that the simplicity of the penalty scheme is indeed reflected in a good understanding of the system: the majority of respondents understand very well the scheme's stepwise shape with its discontinuous changes at cutoffs. This finding is by no means trivial and – potentially due to the simplicity of the penalty structure – sets it apart from a growing body of evidence for individuals' limited knowledge and misperception of non-linear budget sets (e.g., De Bartolome, 1995; Liebman and Zeckhauser; 2004, Chetty et al. 2013; Feldman et al., 2015). The good understanding of the notches observed in our set-up suggests that drivers should respond to the penalty structure. In fact, we asked participants if they would speed on highways and, if at all, whether they would try to avoid higher fines by staying under a certain cutoff. Among those that admitted speeding (almost three of four respondents), 85% indicated they would drive at a speed level slightly below one of the cutoffs in the penalty scheme. Before studying whether we indeed observe this pattern, we now analyze individuals' speeding choices theoretically.

3.4 Theoretical Framework

To set the stage for our empirical analysis, we analyze a risk neutral driver's optimal speeding response to a stepwise penalty scheme. Let the monetary equivalent of the net benefits from a given speed x (time spent on the trip, net of costs for fuel consumption, experienced 'pleasure' from driving at speed x, etc.) be given by a twice differentiable function $v(x,\theta)$, where the parameter θ captures heterogeneous preferences. Drivers' types θ are distributed continuously with density $g(\theta)$ and the c.d.f. $G(\theta)$. For every type θ, $v(.,\theta)$ is concave in x and satisfies the single-crossing property, i.e. $\frac{\partial^2 v(x,\theta)}{\partial x^2} < 0$ and $\frac{\partial^2 v(x,\theta)}{\partial x \partial \theta} > 0 \, \forall x, \theta$.[13] With probability p, the driver's speed is measured by a speed camera. In this case he may get a penalty $f(x)$ which is a step function of the observed speed x:

[13] For a less stylized model of speeding choices (in which, however, drivers' preferences are homogeneous) see Jondrow et al. (1983).

$$f(x) = \begin{cases} f_0 = 0, & \text{if } \; x \leq x_0, \text{ with } x_0 \text{ capturing the speed limit;} \\ ... & \\ f_i, & \text{if } \; x_{i-1} < x \leq x_i, \\ ... & \\ f_I, & \text{if } \; x_{I-1} < x, \end{cases} \tag{1}$$

with x_i denoting the cutoff for speed bracket $i = 1, \ldots, I$, and f_i expressing the costs of the penalties for a given speeding bracket i.[14] A notch at a cutoff speed x_i is given by $\Delta_i := f_{i+1} - f_i > 0$. Assuming that the drivers' utility functions are quasi-linear, their objective functions are given by the net benefits from driving $v(x, \theta)$ and the expected penalties $f(x)$:[15]

$$\max_x EU(x, \theta) = v(x, \theta) - pf(x). \tag{2}$$

As $f(.)$ is a step function which is 'flat' between the different cutoffs x_i, the first-order condition for an *interior* optimum x^* is

$$\frac{\partial v(x^*; \theta)}{\partial x} = 0. \tag{3}$$

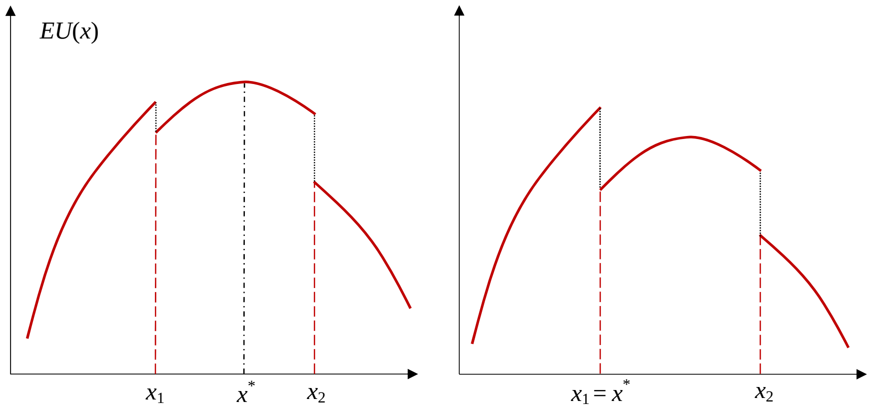
With the additive separable utility function, the interior option is independent of the enforcement parameters. x^* is thus equal to the driver's optimal speeding decision absent of any penalty scheme. Our assumptions on $v(.)$ further imply that $x^* = x^*(\theta)$ is a continuously increasing function of θ.

While the interior solution does not depend on the penalty, the stepwise shape of $f(x)$ gives rise to possible *corner solutions*. Figure 3.2 illustrates this point graphically, plotting x on the horizontal and expected utility (EU) on the vertical axis. The stepwise penalty scheme implies that the inverted-U shape of expected utility discontinuously drops at each cutoff x_i (in the graph: x_1 and x_2). At these speed levels the penalty increases ($\Delta_i > 0$) and, consequentially, the expected utility decreases. As illustrated in the left panel of Figure 3.2, a notch does not necessarily imply a corner solution. Only if the *expected* notch, $p\Delta_i$, is sufficiently large (for a given driver θ), the driver's optimal speed corresponds to a corner solution at a cutoff. This case is depicted in the right panel of the figure.

[14] As the penalty may include non-monetary components (e.g., a driving ban or penalty points), $f(x)$ denotes the *average* present value of the monetary equivalent of the penalty. Allowing for heterogeneity in the penalty across different drivers would complicate the following discussion without yielding additional insights.

[15] Risk aversion would not affect our analysis as long as cross derivatives of Bernoulli utility functions with respect to the net benefits of driving and penalties are zero. If they are not, a driver's interior optimum x^* would depend on her type *and* the size of the penalty.

Figure 3.2: Optimal speed level with notches: Interior optimum and corner solution



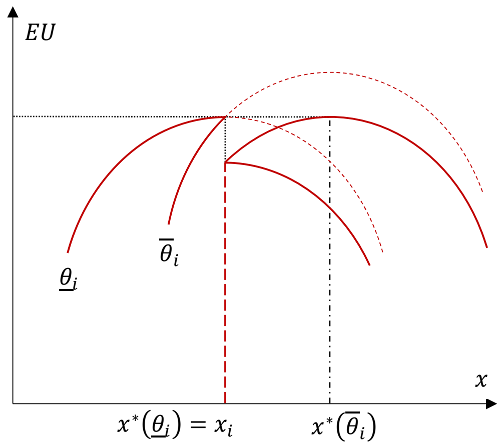
Note: The figure displays the mapping of speed x into expected utility, EU, for a given θ-type and notches at two cutoffs, x_1 and x_2. For the case depicted in the left panel, the expected notch is small and the driver's optimal speed corresponds to the interior solution. In the right panel, the expected notch at x_1 is larger, thus turning this cutoff into the driver's optimal speed level.

3.4.1 Responses to Notches

To study the impact of a penalty scheme with notches more formally, we follow the theoretical analysis in Kleven and Waseem (2013). Note first that, absent of any penalties, the function $x^*(\theta)$ would simply map types of drivers into speed choices. The observed distribution of speed, $H(x)$, would be continuous. In the presence of notches this will generally not be the case. This point is illustrated in Figure 3.3, which plots expected utility for two types who face a notch at x_i. The interior solution for the driver of type $\underline{\theta}_i$ exactly corresponds to the cutoff, $x^*(\underline{\theta}_i) = x_i$, i.e., he would choose the cutoff x_i even absent any penalty scheme. Drivers with slightly higher θ are strictly better off when choosing their corner solution x_i rather than their interior solution $x^*(\theta) > x_i$ which would trigger a penalty $f_{i+1} > f_i$. In contrast, the driver with $\overline{\theta}_i$ is indifferent between his interior solution, $x^*(\overline{\theta}_i) > x_i$ and the corner solution at x_i. For the case depicted in Figure 3.3, all types above $\underline{\theta}_i$ and below $\overline{\theta}_i$ strictly prefer the corner solution x_i over any speed $x > x_i$. The set of drivers who bunch at the cutoff x_i is thus given by the type interval $[\underline{\theta}_i, \overline{\theta}_i]$. The following Lemma generalizes this observation for a penalty scheme with more than one notch:[16]

[16]All proofs are provided in the Appendix.

Figure 3.3: Bunching at notch x_i.



Note: The figure displays the mapping of speed x into expected utility for heterogenous θ-types and one notch at the cutoff x_i. For the driver with type $\theta = \underline{\theta}_i$, the corner solution at the cutoff is identical to her interior solution. The driver with $\theta = \overline{\theta}_i$ is indifferent between her the interior solution $x^*(\overline{\theta}_i)$ and the corner solution at x_i. All drivers with $\underline{\theta}_i < \theta < \overline{\theta}_i$ will prefer the speed x_i over their interior optimum $x^*(\theta)$.

Lemma 1 *Consider a notch $\Delta_i > 0$ at speed cutoff x_i that is used by a non-empty set of types. The probability mass of drivers speeding at x_i is given by $\Pi_i = G(\overline{\theta}_i) - G(\underline{\theta}_i)$, where $\underline{\theta}_i$ satisfies either*

$$x^*(\underline{\theta}_i) = x_i \tag{4}$$

$$\text{or} \quad \exists j < i : v(x_i, \underline{\theta}_i) - v(x_j, \underline{\theta}_i) = p(f_i - f_j), \tag{5}$$

and $\overline{\theta}_i$ satisfies either

$$\exists j > i : v(x^*(\overline{\theta}_i), \overline{\theta}_i) - v(x_i, \overline{\theta}_i) = p(f_j - f_i) \qquad \text{and} \qquad x_{j-1} < x^*(\overline{\theta}_i) < x_j, \tag{6}$$

$$\text{or} \quad \exists j > i : v(x_j, \overline{\theta}_i) - v(x_i, \overline{\theta}_i) = p(f_j - f_i). \tag{7}$$

Lemma 1 shows that there could be different characterizations of the boundaries of the interval $[\underline{\theta}_i, \overline{\theta}_i]$. In Figure 3.3, the boundaries are characterized by (4) and (6) (with $j = i + 1$). In the case of multiple notches, however, the lower bound could also be given by (5). For this case, the type with $\underline{\theta}_i$ would be indifferent between a corner solution at x_i and a cutoff at a lower speed bracket, $x_j < x_i$. Similarly, the upper bound could be characterized by (7), which describes a type who is indifferent between a corner solution at x_i and a corner solution at a cutoff for a higher speed bracket, $x_j > x_i$.

The key implication from Lemma 1 is that notches may push drivers into corner solutions. Empirically, we should thus observe bunching of drivers at a speed level equal to a cutoff, $x = x_i$, and a sparsely populated (or even empty) range of speed levels slightly above a cutoff. The latter 'density holes' in $H(x)$ stem from drivers in the interval $[\underline{\theta}_i, \overline{\theta}_i]$ who would, in the absence of law enforcement (for $p = 0$), choose a speed in the range $x^*(\underline{\theta}_i) < x_i < x^*(\overline{\theta}_i)$. We will discuss the sensitivity of these predictions below.

3.4.2 Responses to the Reform

To assess the impact of the 2009 reform on speeding, let us first consider a simple, hypothetical reform: an increase in f_h by a constant amount for all speed brackets $h > \ell$. Such a reform increases Δ_ℓ at cutoff x_ℓ but leaves all other notches unaffected. All speed levels $x > x_\ell$ become less attractive and drivers will choose a weakly slower speed. In terms of the distribution $H(x)$, some mass of drivers located above x_ℓ before the reform will be shifted towards lower speed levels.

Lemma 2 *Consider a reform that increases one notch Δ_ℓ and leaves all other Δ_j, $j \neq \ell$ unchanged. Then every driver will drive weakly slower after the change.*

The hypothetical reform will also affect bunching. On the one hand, some types of drivers with an interior optimum $x^*(\theta) > x_\ell$ before the reform will start to bunch at cutoff x_ℓ (or a 'lower' cutoff x_l, $l < \ell$). Hence, bunching at x_ℓ (and lower cutoffs) tends to increase. On the other hand, driver types who were initially bunching at a cutoff x_h, $h > \ell$, will find it more attractive to drive slower. Bunching at x_h will then decrease.

Proposition 1 *Consider a reform that increases one notch Δ_ℓ and leaves all other Δ_j, $j \neq \ell$ unchanged. The probability mass Π_i of drivers that bunch at a given cutoff x_i will then*
(i) weakly decrease if $i > \ell$, and (ii) weakly increase if $i \leq \ell$.

In a nutshell, Proposition 1 shows that bunching at a given cutoff x_i increases in the size of a notch above this cutoff. Vice versa, bunching at x_i decreases if a notch at a lower speed cutoff increases. Obviously, the reform described in Table 3.1 differs from our hypothetical case. The 2009 reform was characterized by an increase of the notch at 20km/h above the limit and of all notches at 30km/h and above. The logic behind Proposition 1 thus predicts an increase in bunching at the 20km/h cutoff (and at all lower cutoffs) after the reform. For all other cutoffs, however, the impact of the reform is ambiguous. To see this point, consider the 30km/h cutoff, which experienced the largest increase of the notch. The larger notch at 30km/h (and the increase of 'higher' notches) tends to induce more bunching at this cutoff, whereas the increase of the notch at 20km/h works in the opposite direction. Without further assumptions, the overall effect on the bunching mass at the 30km/h cutoff is therefore unclear. Independently of the bunching, however, our analysis suggests that the reform should result in weakly slower speed levels (see Lemma 2).

3.4.3 Discussion

The analysis from Section 3.4.1 suggests that we should expect bunching at cutoffs together with density holes in the speed range above a cutoff. There are, however, several arguments why this might not be borne out by the data. The most important arguments are based on the difficulties in targeting a specific cutoff x_i. First, there is substantial variation in the speed indicated by speedometers of different automobiles. Hence, a driver who observes a speed of, for instance, 130km/h on the car's speedometer will most likely not drive $x = 130$km/h.[17] This also means that cruise controls (which are fairly uncommon in Germany) do not necessarily facilitate the targeting of cutoffs.

Second, our notation indicates that we model the drivers' choice over a *penalty-relevant* speed x, i.e., the actual speed after applying the tolerance rule (see Section 3.2). Choosing a speed which corresponds to a cutoff x_i thus requires drivers to correctly compute the way in which the tolerance rule maps the measured speed s into the penalty-relevant speed x. While the survey evidence suggests that roughly every third driver exactly knows the tolerance rule, two thirds either over- or underestimate the rule's generosity. As a result there might be a significant amount of optimization errors – which are, presumably, more frequent than in the context of tax notches (see Chetty, 2012; Kleven and Waseem, 2013). These errors will work against bunching and will diminish any density holes.

A further and important reason why we might not see much bunching is based on the possibility that drivers – particularly those who choose to drive above the speed limit – underestimate the risk of a speed control. This derives from the fact that,

[17] The European Council Directive 75/443/EEC and §57(2) of the German *Straßenverkehrs-Zulassungsordnung* allows a tolerance in the speed displayed by speedometer of up to 13% *above* the true speed level (for the relevant speed range studied below). An indicated speed of 130km/h could therefore correspond to an actual speed of only 115km/h.

cet. par., bunching will be proportional to the *expected* notch, $p\Delta_i$. The lower the drivers' prior about the probability p, the more likely they are to choose their interior optima. If most speeders are driving above the speed limit because they underestimate the detection risk, we should therefore observe no bunching. However, if there is a second type of speeders who drive above the limit despite a (sufficiently) high prior about p, these types will be bunching. In combination, heterogeneous priors might produce bunching without any pronounced density holes.

Related to the individuals' risk assessments, there is evidence from neuroscience suggesting that drivers display a diminished control of risk-taking behavior at higher speed levels – when the control of the car requires more cognitive resources (Jäncke et al., 2008). In our context this would imply that drivers are more likely to act 'as if $p \approx 0$' at a higher speed. As a consequence, we should observe less bunching at higher speed bracket cutoffs. Our rational choice model would yield an equivalent prediction, if the 'taste for speeding' (captured by the curvature of $v(x, \theta)$) would become more 'sharp' for high values of θ (as reflected in $\partial^3 v(x, \theta)/\partial x^2 \partial \theta$). We will return to these arguments below.

A last point worth discussing is the fact that our analysis – in contrast to the taxation literature (e.g., Saez, 2010; Chetty et al., 2011; Kleven and Waseem, 2013) – does *not* link the bunching mass to the elasticity of speeders with respect to the fine. There are two important reasons for not performing such an analysis. First, agents respond to two dimensions of policy: the penalty scheme (reflected in $f(x)$) and the detection probability p. While information on the scheme are available to drivers (just like the tax rates and thresholds are, in principle, observable for taxpayers), probabilities are largely unknown. In the context of heterogeneous priors one would then need very strong assumptions to identify the relevant elasticity.[18] Second, the type range of drivers who bunch in our context might be given by equations (5) and (7) from Lemma 1 – a case which is neglected in tax studies, because tax thresholds are typically located in quite different (e.g., income) ranges. Our set-up, however, is characterized by many 'closely' located notches. Technically, we would need to identify both the highest and the lowest types of drivers bunching at a given notch, which implies one additional identification requirement.

[18]Note further that the elasticities which derive from bunching estimates are sensitive to the specific functional form of preferences. While recent taxation research is characterized by an (implicit) consensus about the 'right' utility function, we are not aware of any consensus on drivers' preferences over speeding and monetary well-being.

3.5 Data

To empirically evaluate speeding behavior, we use data from a highway police unit which is responsible for monitoring 575 kilometers of Autobahn in the state of North Rhine-Westphalia. The police unit provided us with information on *all* tickets that emerged from speed controls with mobile cameras on stretches of the Autobahn with a speed limit of 100km/h during the period 01/2005 to 12/2006 and 01/2009 to 03/2010. The data cover 154,970 speeding tickets with an overall amount of 10.85m Euros in fines. For each ticket we observe the penalty-relevant speed x in integer values (i.e., the outcome from applying the tolerance rule discussed in Section 3.2), the precise date, time and location of the speed measurement as well as the weather (sunny, cloudy, rainy) and street conditions (dry, damp, wet). For a subset of the tickets, we also observe the driver's gender and several digits of the car's license plate. The latter information allows us to identify local drivers.[19]

Table 3.3, which provides summary statistics on our data, indicates that around 70% of the observations come from the pre-reform period. The data cover speeding tickets from all days of the week, with fewer tickets on weekends and slightly more on Wednesdays. More than 40% of the speeding offenses were recorded in the morning (8am–12am), around 25% in the evening (4pm–8pm) and less than 5% at night (8pm–12pm). For the sub-sample of tickets with richer information we find that around 80% of speeders are male and roughly 20–30% are locals.

The data further reveal that not every speed measurement with $x > 100$km/h resulted in a ticket. In 86% of all speed control sessions (covering 83% of all speeding tickets), the police only recorded and enforced speeding offenses with $x \geq 116$km/h.[20] Hence, we only observe the truncated distribution of penalty-relevant speed measures. Among these measures, the average speed x is 125.11 km/h, with a slightly lower speed in the post- as compared to the pre-reform period – an observation that we will explore in more detail below. Finally, the average fine is 63.82 euro in the pre-reform sample. After the reform, this value increases by 35% to 86.29 euro.

[19] A driver is coded as local if the *Kreis* (county) indicated by the license plate corresponds to the 'home' or a neighboring Kreis of the location of the speed measurement.

[20] This practice is a response to the administrative costs of issuing and enforcing a ticket. These costs make speeding tickets which include only small fines economically unattractive. Our data also show, however, that the police sometimes enforce minor speeding violations (starting with 106km/h).

Table 3.3: Summary statistics – speeding tickets

	Pooled data		*Pre-reform*		*Post-reform*	
Variable	Mean	(Std. Dev.)	Mean	(Std. Dev.)	Mean	(Std. Dev.)
Speeding x	125.11	(9.20)	125.16	(9.04)	124.99	(9.60)
Monetary fine f	70.23	(49.72)	63.82	(39.01)	86.29	(67.02)
Enforcement limit:						
= 116km/h	0.83	(0.38)	0.82	(0.39)	0.86	(0.35)
= 121km/h	0.10	(0.30)	0.12	(0.33)	0.03	(0.18)
Male Drivers*	0.83	(0.37)	0.84	(0.36)	0.80	(0.40)
Local Drivers*	0.27	(0.44)	0.32	(0.47)	0.17	(0.37)
12:00 am – 7:59 am	0.00	(0.02)	0.00	(0.01)	0.00	(0.03)
8:00 am – 11:59 am	0.43	(0.50)	0.46	(0.50)	0.38	(0.48)
12:00 pm – 3:59 pm	0.28	(0.45)	0.24	(0.43)	0.36	(0.48)
4:00 pm – 7:59 pm	0.24	(0.43)	0.25	(0.43)	0.21	(0.40)
8:00 pm – 11:59 pm	0.05	(0.22)	0.04	(0.20)	0.06	(0.24)
January	0.09	(0.28)	0.11	(0.31)	0.04	(0.19)
February	0.07	(0.26)	0.06	(0.24)	0.11	(0.31)
March	0.10	(0.30)	0.09	(0.29)	0.11	(0.32)
April	0.09	(0.28)	0.07	(0.26)	0.13	(0.33)
May	0.11	(0.32)	0.11	(0.31)	0.12	(0.32)
June	0.08	(0.27)	0.07	(0.25)	0.11	(0.31)
July	0.07	(0.26)	0.08	(0.27)	0.06	(0.24)
August	0.09	(0.28)	0.09	(0.29)	0.07	(0.26)
September	0.11	(0.31)	0.12	(0.33)	0.07	(0.25)
October	0.09	(0.29)	0.11	(0.31)	0.06	(0.24)
November	0.06	(0.23)	0.06	(0.24)	0.05	(0.23)
December	0.04	(0.20)	0.03	(0.17)	0.07	(0.25)
Monday	0.16	(0.37)	0.16	(0.37)	0.15	(0.36)
Tuesday	0.15	(0.36)	0.16	(0.37)	0.12	(0.32)
Wednesday	0.21	(0.40)	0.20	(0.40)	0.23	(0.42)
Thursday	0.15	(0.36)	0.15	(0.36)	0.16	(0.37)
Friday	0.14	(0.34)	0.13	(0.34)	0.15	(0.35)
Saturday	0.09	(0.28)	0.08	(0.27)	0.09	(0.29)
Sunday	0.11	(0.31)	0.11	(0.31)	0.11	(0.31)
Number of:						
speeding tickets	154,970		110,721		44,249	
speed control sessions	1,139		843		296	

Notes: The table presents summary statistics – sample means and standard deviations in parenthesis – on the speeding tickets from the pooled, the pre- and the post-reform sample. * indicates that the variable is only recorded in a sub-sample of tickets. The small share of observations in post-reform January is due to the fact that the reform was introduced on 1 February 2009.

3.6 Results

3.6.1 Descriptive Evidence

We start out by examining whether the distribution of x among the speeding tickets provides any evidence for bunching at the cutoffs of the penalty scheme. Figure 3.4(a) illustrates the density distribution of the penalty-relevant speed x among the pooled sample of all speeding tickets with $x \geq 116$ relative to the total population of roughly 6 mio. measured drivers. The dashed green lines indicate the cutoffs from the penalty function (see Table 3.1). The density distribution is decreasing in the speed level and displays several major spikes. Two of these spikes are located right at cutoffs (120 and 125km/h), and one is located slightly below a cutoff (129 km/h). The figure does not show any pronounced density holes 'to the right' of the spikes – a point that we will return to below. For speed levels with $x \leq 122$, we observe a lot of variation in the density distribution.[21] This makes it hard to evaluate the spike at the lowest cutoff (120 km/h). While it is more clear that there is no visible evidence for bunching at higher speed cutoffs (140, 150 and 160km/h), the distribution shows several drops, some of which overlap with the cutoffs.

Recall that our data capture the penalty-relevant speed x *after* applying the tolerance rule (see above). Note further that the way in which the tolerance rule transforms the measured speed s into the penalty-relevant speed x mechanically produces a concentration of tickets at some values of x. In particular, all measures with $133 \leq s < 135$ [$166 \leq s < 168$] will be recorded with $x = 129$ [$x = 161$] in our data. Hence, the spike at $x = 129$ [and at $x = 161$] which is illustrated in Figure 3.4(a) might be a result of the tolerance rule's non-injective mapping of s into x.[22]

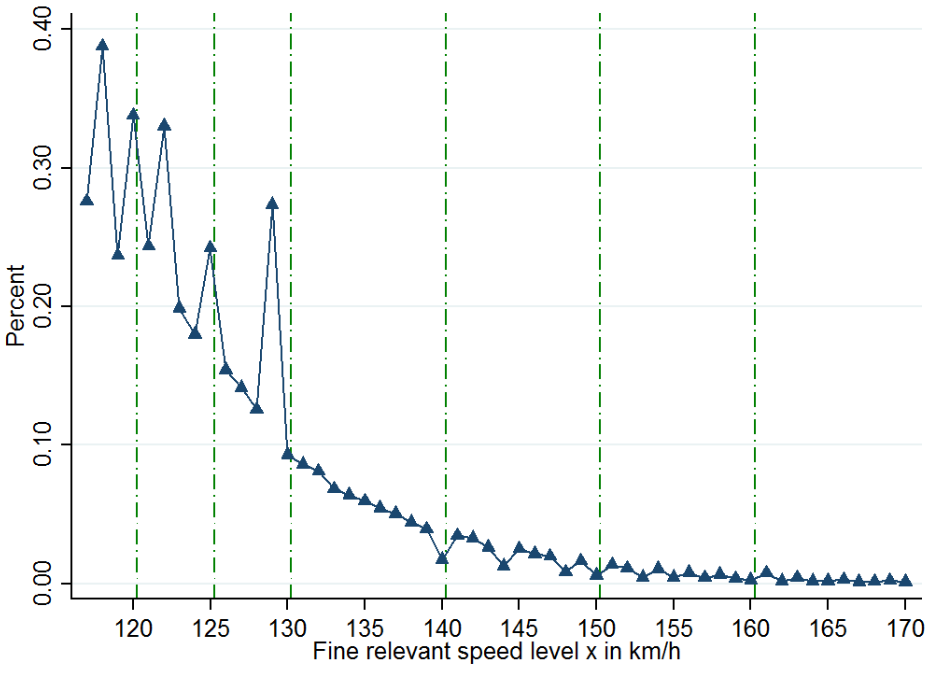
To account for this mechanical effect, we empirically 'revert' the mapping. For each value of x which maps one-to-one into s, we first assign the correct speed measured s (in integer values). Omitting the values for $s = \{133, 134, 166, 167\}$ we then estimate a higher-order polynomial function that approximates the observed distribution of speed tickets over s. Based on the estimated distribution we finally assign the density mass

[21]We discussed this observation with the police unit which provided us with the data. While we could not identify any plausible explanation for the variation in the density for speed $x \leq 122$, we can reject the hypotheses that the variation is induced by different measurement techniques, rounding issues, or by overlapping enforcement thresholds.

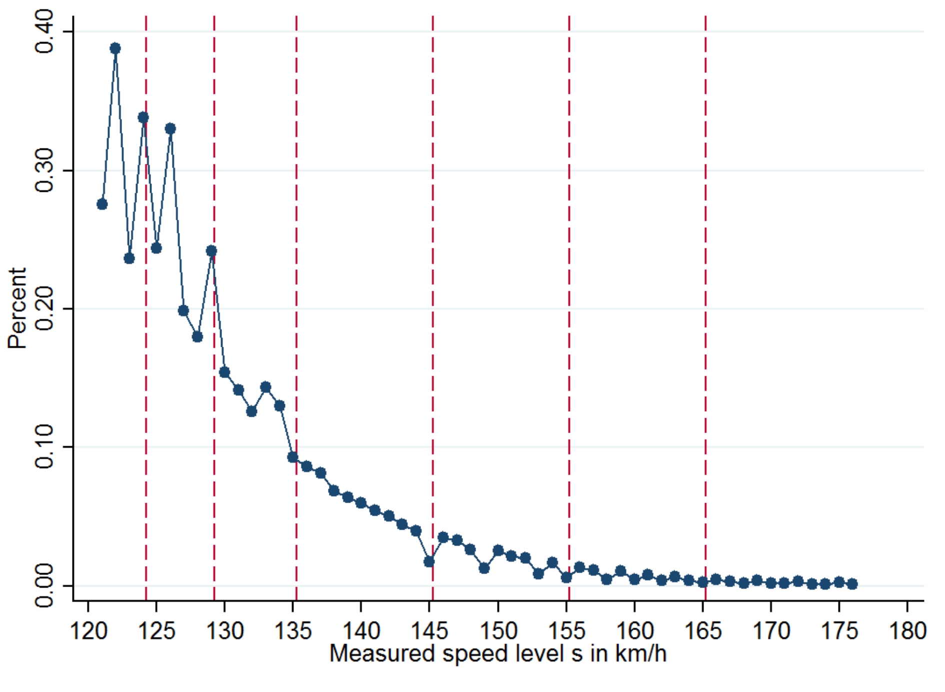
[22]To illustrate the problem, note that for any $133 \leq s < 135$ the tolerance rule – rounding down, subtracting 3% of the speed and rounding down again – will transform s into $x = 129$. For $135 \leq x \leq 165$, however, the tolerance rule is bijective, mapping *one* value of measured speed s into *one* observed value of x.

Figure 3.4: Density distribution of speed

(a) Observed distribution of penalty-relevant speed x



(b) Projected distribution of measured speed s



Notes: The figure illustrates the observed distribution of the penalty-relevant speed x (Panel a) as well as the projected distribution of measured speed s (*Panel b*) among all tickets (pooling data from the pre- and post-reform period). The vertical axes indicates the fraction of tickets observed for a given speed level, relative to the total number of speed measures. The horizontal axes capture the penalty-relevant speed x (Panel a) and the measured speed s (Panel b), respectively. The speed limit is 100km/h. The dashed green lines indicate the cutoffs x_i from the penalty function (Panel a); the dashed red lines (Panel b) express these cutoffs in terms of the measured speed s.

from $x = \{129\}$ $[x = \{161\}]$ to the speed levels $s = \{133, 134\}$ $[s = \{166, 167\}]$. The resulting distribution is presented in Figure 3.4(b). The figure shows that the massive spike at $x = 129$ from Figure 3.4(a) considerably shrinks once we account for the rounding rule. Nevertheless, the projected distribution includes a significant heap at $s = \{133, 134\}$ (corresponding to $x = 129$).[23] Hence, the spike below the 130km/h cutoff is only partially due to the rounding rule.

3.6.2 Estimation Approach

To estimate the bunching mass at a cutoff x_i, we start from the empirically observed mass of tickets within the range $[s(x_i) - \delta; s(x_i)]$, where $s(x_i)$ indicates a cutoff from the penalty scheme (in terms of measured speed s) and $\delta \geq 0$ defines the bunching area below the cutoff (in integer km/h values of measured speed). Following the taxation literature, we then assess this mass of speeders relative to the expected mass from a counterfactual distribution for the hypothetical case without a notched penalty scheme.

To obtain the counterfactual we approximate the speed ticket distribution from Figure 3.4.b by a polynomial function. More specifically, we estimate

$$C_s = \sum_{q=0}^{\bar{q}} \beta_q \, s^q + \sum_{r=(s(x_i)-\delta)}^{s(x_i)} \gamma_r \cdot \mathbf{I}[s = r] + \epsilon_s \tag{8}$$

where C_s indicates the share of drivers measured with speed s, $\bar{q}$ defines the order of the polynomial function, and $\mathbf{I}$ is the indicator function. Based on the estimated β-coefficients (but excluding the γ-coefficients) we then predict $\hat{C}_s^{prox} = \sum_{q=0}^{\bar{q}} \hat{\beta}_q \, s^q$. This initial proxy for the counterfactual distribution neglects the excess mass of speeders from the range $[s(x_i) - \delta; s(x_i)]$ who would, in the absence of a notched penalty scheme, choose a speed level 'to the right' of the cutoff $s(x_i)$. To account for this fact, $\hat{C}_s^{prox}$ is inflated for speed values $s > s(x_i)$, up to the point where the counterfactual distribution of $\hat{C}_s$ satisfies the integration constraint, $\sum \hat{C}_s = \sum C_s$ (i.e., when the empirical and the counterfactual distribution cover an equal number of speeding tickets). The bunching

[23] Note that our approach assumes that the distribution of the observations from one speed level x among the two speed levels in s follows the estimated, 'smoothed' distribution. Hence, the projection ignores that rational drivers could anticipate the property of the tolerance rule and locate predominantly at the higher of the two speed levels s. If drivers indeed behave like this, the 'excess mass' would concentrate on $s = 134$. The following analysis will account for the fact that we cannot determine the precise speed measure for the two pairs of s.

mass $\hat{b}_i$ in the speed range $[s(x_i) - \delta; s(x_i)]$ is then given by

$$\hat{b}_i = \sum_{s=(s(x_i)-\delta)}^{s(x_i)} \frac{C_s - \hat{C}_s}{\hat{C}_s/(1+\delta)}. \tag{9}$$

$\hat{b}_i$ indicates the excess mass, i.e., the difference between the observed and the predicted speed tickets with $s(x_i) - \delta \leq s \leq s(x_i)$ (in the numerator) relative to the average mass in the counterfactual distribution for this range (denominator). We estimate $\hat{b}_i$ together with boot-strapped standard errors using the iterative procedure from Chetty et al. (2011).

Several aspects of our approach deserve a closer discussion. Note first that we base our estimates on the projected distribution of speed s rather than the distribution of penalty-relevant speed values x. By doing so, and by accordingly adjusting δ, we avoid potential problems with the rounding rule. Second, we will report bunching estimates that *locally* approximate the counterfactual distribution for the speed range around each cutoff. Our results remain qualitatively unaffected when we estimate one counterfactual for the full range of s. The same holds true when we estimate bunching using simpler approximations (e.g., in the spirit of Saez, 2010).

Last, the estimation of (8) accounts for the observations from the bunching area $[s(x_i) - \delta; s(x_i)]$ but not for those in the range 'to the right' of a cutoff, with a potential missing mass in the distribution (compare Kleven and Waseem, 2013). This approach, which is closer to a 'kink'- rather than a 'notch-bunching' analysis, is motivated by the fact that we face a high number of nearby notches with only few (integer valued) observations between two notches. This prevents us from jointly estimating a bunching and a missing-mass area (as in Kleven and Waseem, 2013). Moreover, we do not observe any pronounced density holes (see Figure 3.4). Note further, that we will not use the bunching mass estimates to compute a proxy for a behavioral elasticity (see the discussion in Section 3.4.3).

3.6.3 Bunching Estimates

Figure 3.5(a) and (b) present the results from bunching estimates for the cutoffs with 25 and 30km/h above the limit, respectively. For the moment, we pool the data for the pre- and post-reform period. Figure 3.5(a) shows sharp bunching right at the cutoff $s(x_i) = 129$. The estimated bunching coefficient for the range $s = \{128, 129\}$ (i.e., at

$x_i = 125$ with $\delta = 1$) indicates an economically and statistically significant excess mass of 36% (relative to the average counterfactual mass in that speed range). Figure 3.5(b), which presents the estimate for the cutoff at $s(x_i) = 135$, illustrates an excess mass which is located at least one km/h below the cutoff. To account for the fact that we cannot distinguish the measured speed for $s = 133$ and $s = 134$ (see Section 3.6.1), we estimate bunching in the broader range $s = \{133, 134, 135\}$ (i.e., we set $\delta = 2$ for $x_i = 130$).[24] Just like for the first cutoff, we obtain a significant bunching mass of 42%.

What about bunching at the other cutoffs of the penalty scheme? Recall first that the majority of the speeding tickets are based on violations with $x \geq 116$. Hence, we cannot study bunching at the first two cutoffs, at 110 and 115km/h. As discussed above, there is a sizeable spike together with a substantial amount of unexplained variation in the distribution of tickets around the cutoff at $x_i = 120$km/h (compare Figure 3.4). Our method to quantify bunching thus yields a positive but imprecisely estimated coefficient for this cutoff (see Table 3.4).

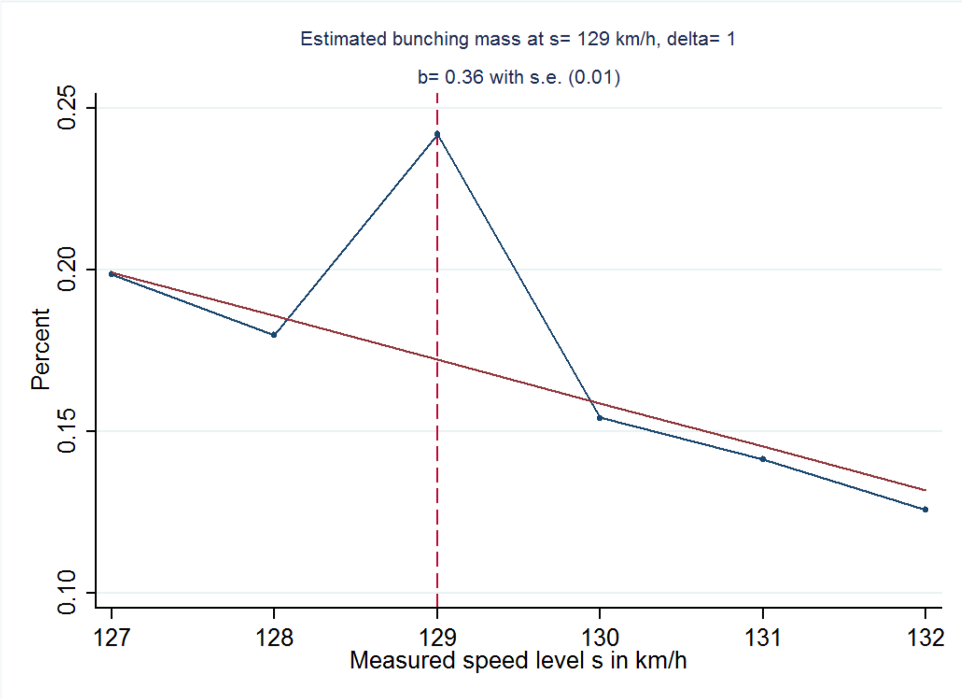
Consistent with the graphical analysis from above, we do not detect any evidence on bunching for cutoffs at higher speed levels (see Table 3.4).[25] The null-results for these high speed cutoffs – which would correspond to an actually measured speed of $s = 145$, 155 and 165km/h, respectively – are consistent with the evidence showing that drivers tend to make less deliberate choices when they drive at very high do speed: their cognitive capacities are depleted which in turn reduces the capacity to optimally trade-off risks (Jäncke et al., 2008). Two alternative explanations might be that rational drivers, who drive 40 and more above a speed limit of 100km/h, either have a very sharp 'taste for speeding' (as captured by $\partial^3 v(x, \theta)/\partial x^2 \partial\theta > 0$) or they expect a very low detection risk p. In either case, they would be fairly insensitive to the notches in the penalty scheme.

[24] With $\delta = 2$, the estimated coefficient is insensitive to how the projection allocates speed tickets from $x = 129$ to $s = 133$ and $s = 134$ (see Section 3.6.2).

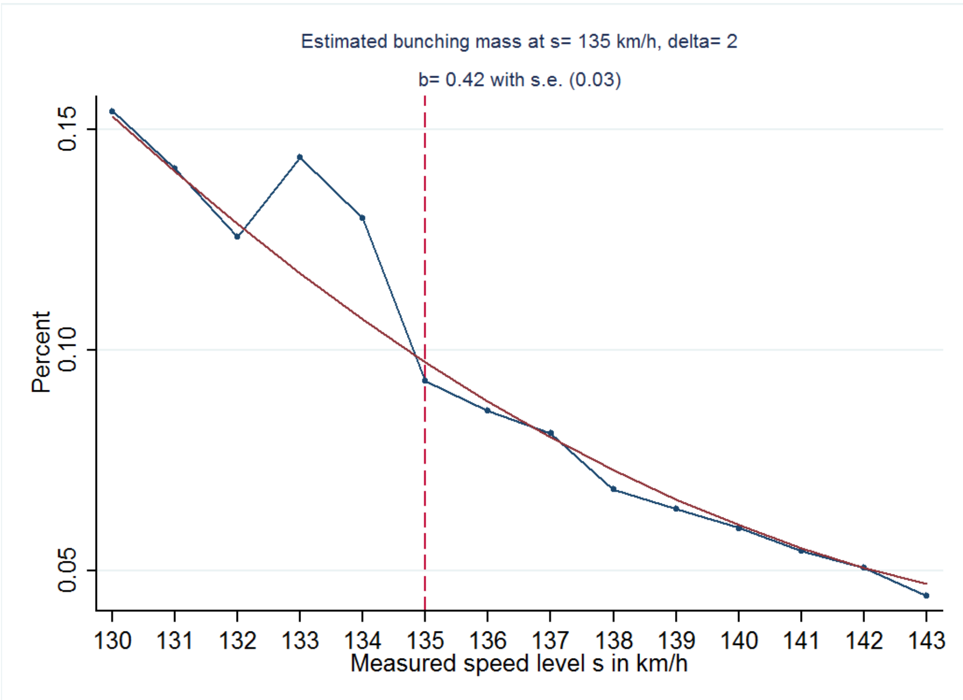
[25] For the cutoff at $x_i = 140$km/h, we estimate a significantly negative value for $\hat{b}_i$. Note, however, that we observe a fairly small number of tickets together with a substantial variance in their distribution in this speed range. It is therefore not surprising that for high speed values the estimates for $\hat{b}_i$ are sensitive to the choice of the polynomial degree.

Figure 3.5: Empirical and counterfactual distribution of speeding levels

(a) Bunching estimation at the 125km/h cutoff



(b) Bunching estimation at the 130km/h cutoff



Note: Empirical and counterfactual distribution of measured speeding for a speed limit of 100km/h from pre- and post-reform data (pooled). The counterfactual distributions for graphs (a) and (b) are based on a linear (i.e., $\bar{q} = 1$) and a quadratic fit ($\bar{q} = 2$ for $s(x_i) = 135$), respectively. The horizontal axes indicates the empirical speed above the limit. The vertical axes indicates the percentage share of observations for each speed level (relative to all measured drivers). The red dashed vertical line in the *top* graph indicates the speed $s = 129$km/h (corresponding to a penalty-relevant speed $x = 125$km/h). The red dashed vertical line in the *bottom* graph indicates the speed of $s = 135$km/h (corresponding to $x = 130$km/h).

Table 3.4: Bunching estimates for different cutoffs

Cutoff & Bunching Range	*Pooled data*	*Pre-reform*	*Post-reform*
$x_i = 120$ with $\delta = 1$	-0.04	-0.07	0.03
($s = \{123, 124\}$)	(0.25)	(0.24)	(0.24)
$x_i = 125$ with $\delta = 1$	0.36	0.39	0.26
($s = \{128, 129\}$)	(0.01)	(0.02)	(0.03)
$x_i = 130$ with $\delta = 2$	0.42	0.43	0.40
($s = \{133, 134, 135\}$)	(0.03)	(0.04)	(0.08)
$x_i = 140$ with $\delta = 1$	-0.54	-0.55	-0.51
($s = \{144, 145\}$)	(0.15)	(0.13)	(0.19)
$x_i = 150$ with $\delta = 1$	-0.20	-0.30	0.06
($s = \{154, 155\}$)	(0.56)	(0.57)	(0.62)
$x_i = 160$ with $\delta = 1$	-0.50	-0.57	-0.35
($s = \{164, 165\}$)	(0.71)	(0.72)	(0.67)

Notes: The table displays the bunching estimates $\hat{b}_i$ for the cutoffs analyzed in Figures 3.5 and 3.7 as well as for other cutoffs. x_i indicates the cutoff (in terms of penalty-relevant speed), s is the measured speed and δ captures the width of the bunching area.

We conducted several robustness checks and refinements. The bunching estimators for the speed cutoffs at $x_i = 125$ and 130km/h turn out to be highly robust to using alternative specifications (e.g., higher order polynomials) in the approximation of the counterfactual distribution. (For 'higher cutoffs', this is not the case; here we do observe much more variation in the estimated $\hat{b}_i$.) Splitting the sample for different hours of the day, different weekdays or seasons, we detect no significant differences in bunching (or stable null-results). We also split the sample according to different levels of traffic density (approximated by the number of measured cars per hour) and differentiated local and non-local drivers. Again, the data do not indicate any significant differences in bunching behavior. Overall, the estimated bunching masses at $x = 125$ and 130km/h seem to be very robust with respect to other observable characteristics.

In a next step, we estimated the probability of speed controls occurring at a given time (hour, day, month) and stretch of the Autobahn. This allows us to distinguish between 'unlikely' (or surprising) and 'likely' (or expectable) speed controls. The results indicate that bunching tends to increase in the (predicted) probability of a speed control. Similar to the results from our other split-sample exercises, however, the differences are modest. What is much more pronounced, however, is an overall adjustment in the drivers' behavior: when speed controls are more likely to occur, we observe considerably fewer violations of the speed limit in the first place. The exercise therefore suggests that the total population of drivers clearly responds to the variation in the objective detection risk, whereas the (self-)selected sample of speeders is fairly insensitive to the risk – potentially, because they underestimated p.

To sum up, we find evidence for bunching only at two notches of the penalty scheme. In the light of the (i) difficulties in targeting the 'right' speed level (optimization errors), (ii) imperfect knowledge about the tolerance rule and (iii) the impact from underestimating the detection risk, the evidence provides reasonable support for the theoretical prediction of bunching. A striking difference to the literature on tax notches, however, is the absence of density holes. As discussed above, this might be explained by heterogeneous beliefs about p. On the one hand, drivers who anticipate a sufficiently high detection risk but nevertheless decide to speed might choose their optimal speed only among different cutoffs. For these drivers, the *expected* notches, $p\Delta_i$, would be large and a corner solution would dominate all interior solutions. Speed levels between two cutoffs x_i and x_{i+1} would only be observed due to optimization errors. On the other hand, drivers who believe that a speed control is unlikely to occur would choose their interior optima, which are smoothly distributed all over the speed range. The combination of drivers with heterogeneous beliefs could then produce some bunching without having any density holes in the distribution of speeding tickets.

3.6.4 Responses to the Reform

Let us now analyze the impact of the reform. The descriptive statistics from above revealed that, among the speeding tickets, the average speed x declined from 125.16 to 124.99km/h (see Table 3.3). These numbers, however, might be driven by differences in the sample period, i.e., when and under which conditions the speed was measured. To mitigate this issue, we ran a propensity score matching exercise to arrive at a pre- and post-reform sample which is comparable regarding the time (hour of day, weekday, month), enforcement limit, street and weather conditions. The results from this matching exercise are presented in Table 3.5. Similar to the basic descriptives, the results indicate that the reform is associated with a modest 0.23% decline in the penalty-relevant speed x, from 125.36 to 125.07km/h.

Table 3.5: Propensity score matching of pre- and post-reform speeding Levels

Sample	Mean	(Std. Dev.)	1^{st} Quartile	Median	3^{rd} Quartile	N
Pre-reform	125.33	(109.16)	119	123	129	90,486
Post-reform	125.09	(109.69)	118	122	129	32,315

Notes: The table reports the comparison of pre- and post-reform speeding levels from a propensity matching exercise with the following confounders: time (hour of day, weekday, month), enforcement limits, location of speed control, street and weather conditions.

It turns out, however, that Table 3.5 – which only indicates that the average speed *among* speeding tickets is lower in the post-reform period – gives a misleading picture on the impact of the reform. This point becomes obvious once we analyze the change in the speed distribution beyond the selected sample of speeders. To do so, we computed the share of speeding tickets with a recorded speed x above a given cutoff x_i, relative to the total number of cars (speeding and non-speeding) which were measured during each speed control session. The results from this exercise are presented in Table 3.6.

Table 3.6: Fraction of speeders relative to *all* measured cars

Speed measure with...	*Pre-reform*	*Post-reform*
$x > 120$km/h	0.030	0.023
	(0.028)	(0.021)
$x > 125$km/h	0.019	0.014
	(0.020)	(0.012)
$x > 130$km/h	0.011	0.008
	(0.013)	(0.008)
$x > 140$km/h	0.004	0.003
	(0.006)	(0.003)
N (speed control sessions)	843	296

Notes: The table presents the fraction of speeding tickets with a speed x above different cutoffs from the penalty function, relative to all (speeding and non-speeding) cars measured per speed control session. Standard deviations are in parenthesis.

The table reveals a pronounced shift in the speed distribution: relative to all drives, the fraction of speeders with $x > 120$km/h – i.e., cars driving in the speed range for which the reform increased the fines (see Table 3.1) – dropped from 3.0 to 2.3%. Considering the cutoffs x_i at 125, 130 and 140km/h, we observe a similarly strong decline of roughly 25% in the fraction of speeders who got ticketed with $x > x_i$. While this pre- and post-reform comparison is sensitive to other time-varying factors beyond the reform, the pattern from Table 3.1 is again confirmed by propensity score matching.[26] Hence, the data are consistent with the prediction from Lemma 2: increasing the fines at $x_i = 120$km/h (and 'higher' cutoffs) renders speeding in this range less attractive. We observe a pronounced shift in the speeding distribution with the fraction of speeding tickets with $x > 120$km/h dropping by 23%.

In a next step we study whether bunching at cutoffs changed between the pre- and the post-reform period. Graphical evidence suggests that bunching at the two cutoffs $x_i = 125$ and 130km/h is equally observed in the pre- and the post-reform sample (see Figure 3.6 in the Appendix). To assess the changes in the speed distribution, we first consider a simple estimation framework. We estimate the equation

$$\text{Bunching}_j = \mu_0 + \sum_{\ell} \lambda_\ell \left(\text{Reform}_j \times I_j^\ell\right) + \sum_{\ell} \mu_\ell \, I_j^\ell + \mathbf{X}_j \kappa + \varepsilon_j, \tag{10}$$

where Bunching_j is a dummy indicating whether a ticket j with speed x_j falls into the range at or slightly *below* a given cutoff x_ℓ, $x_\ell - \delta \leq x_j \leq x_\ell$. Reform_j indicates whether the ticket is from the post-reform period, I_j^ℓ captures if a speed ticket with x_j is located around a given cutoff, $x_\ell - \delta \leq x_j < x_\ell + \delta$, and $\mathbf{X}_j$ is a vector of control variables (including dummies for the hour, day, month, as well as street and weather conditions during the speed measurement).

For each cutoff ℓ, the coefficients μ_ℓ then captures the fraction of tickets (among those in $x_\ell - \delta \leq x_j < x_\ell + \delta$) that are located at or slightly below x_ℓ. Hence, the μ-coefficients will not capture bunching; they solely reflect the local slope of the (pre-reform) distribution around each cutoff (as captured in Figure 3.6). The coefficients of interest are the λ's, which indicate how the fraction of tickets at or slightly below a cutoff changed after the reform. Linear probability model (LPM) estimates of the λ's from equation (10) are presented in Table 3.7.[27]

Consistent with our theoretical prediction on the impact of the reform, we observe an increase in the mass of tickets at or slightly below the cutoff at 120km/h. The estimates from Table 3.7 suggest that after the reform there is a 5 percentage point higher chance of seeing a ticket with a penalty-relevant speed just below 120km/h in the data. While our theoretical framework does not offer any clear predictions regarding the reform's impact on bunching at other notches (see Section 3.4.2), it is interesting to

[26] Results are available from the authors upon request.

[27] Estimates using non-linear models, which are available from the authors upon request, yield almost identical results.

note that the estimates point to a decline in the frequency of tickets below the cutoffs at 125 and 130km/h – the two cases for which we found strong and robust bunching evidence above. There also seems to be a decline in the mass of tickets below the cutoff at 140km/h, however, the estimate is sensitive to the precise specification and not robust when we vary δ. For the other cutoffs, the regression analysis does not indicate any significant impact of the reform.

Table 3.7: Impact of reform on bunching

	(1)	(2)	(3)	(4)	(5)	(6)
Reform × I^{120}	0.047***	0.050***	0.057***	0.047***	0.047***	0.054***
	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)
Reform × I^{125}	-0.014**	-0.022***	-0.023***	-0.020***	-0.027***	-0.027***
	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.008)
Reform × I^{130}	-0.013**	-0.024***	-0.026***	-0.017**	-0.027***	-0.026***
	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)	(0.009)
Reform × I^{140}	-0.025*	-0.038***	-0.037**	-0.013	-0.024	-0.030
	(0.014)	(0.014)	(0.016)	(0.017)	(0.017)	(0.019)
Reform × I^{150}	0.013	-0.005	-0.014	0.023	0.007	0.008
	(0.025)	(0.025)	(0.028)	(0.027)	(0.027)	(0.030)
Reform × I^{160}	0.062	0.036	0.007	0.066	0.045	0.021
	(0.039)	(0.039)	(0.044)	(0.047)	(0.047)	(0.051)
Control variables:	No	Yes[a]	Yes[b]	No	Yes[a]	Yes[b]
N	154,970	154,970	128,644	154,970	154,970	128,644
R^2	0.665	0.668	0.668	0.571	0.574	0.573

Note: The table presents the outcome from LPM estimates of equation (10). All specifications include (non-interacted) cutoff specific dummies I_j^ℓ. In columns (2) and (5), we control for the year and the enforcement limit of the speed control session. Columns (3) and (6) add further control variables (for the weather conditions, location, month, day of the week and hour of the day). Columns (1)–(3) are based on $\delta = 2$, i.e., we set $I_j^\ell = 1$ if $x_\ell - 2 \leq x_j < x_\ell + 2$. Columns (4)–(6) employ $\underline{\delta} = 1$ and $\overline{\delta} = 2$ and thus set $I_j^\ell = 1$ if $x_\ell - 1 \leq x_j < x_\ell + 2$. The bunching dummies are adjusted accordingly. Robust standard errors are in parenthesis.

To further assess the change in bunching, we also estimated the coefficient $\hat{b}_i$ from equation (9) for the pre- and post-reform period. The results, which are presented in columns 2 and 3 of Table 3.4, capture again an increase in the mass of tickets right at the 120km/h cutoff.[28] For the 125 and 130km/h cutoffs, we observe a decline in bunching. For the former cutoff, the estimated excess mass drops from 39 to 26%; for the latter cutoff we estimate a more modest decline, from 43 to 40% (see Table 3.4 and Figure 3.7 in the Appendix).

[28]Note that the bunching estimates use $\delta = 1$. If we set $\delta = 2$, as in the LPM, the bunching estimates show a more pronounced increase in the excess mass at the 120km/h cutoff from 9 to 36%.

To wrap up, the second part of our empirical analysis points to a non-trivial impact of the reform in 2009, which considerably increased all notches starting with 20km/h above the speed limit (see Table 3.1). In line with theoretical predictions, we observe a 25% drop in the fraction of drivers speeding with 120km/h or above. At the same time, the fraction of speeding tickets slightly below the 120km/h cutoff increases in the post-reform period. Consistent with this pronounced shift in the speed distribution, aggregated accident statistics indicate a positive impact of the reform, too. Comparing the first six months after the reform with the same months in the pre-reform year, the total number of accidents as well as the rate of deadly accidents both declined by 3 percent.[29]

3.7 Concluding Discussion

This paper has studied drivers' knowledge of and responses to a notched penalty scheme for speeders in Germany. We first ran an online survey which provided evidence suggesting that most drivers have a very good knowledge of the scheme's stepwise shape with its discontinuous jumps in penalties at certain speed cutoffs. Exploiting micro-data from more than 150,000 speeding tickets from the German Autobahn, we then studied whether drivers bunch at speed levels slightly below these cutoffs. Consistent with our theoretical analysis, we observe significant bunching at two prominent notches of the scheme. For the notches at very high speed levels (with 40, 50 or 60km/h above the limit of 100km/h), however, there is no bunching. The latter observation is consistent with the interpretation that excessive speeders might have underestimated the risk of a speed control.

This point also highlights one major difference between our analysis and the bunching studies in the taxation literature. In our context, agents respond to *expected notches* which are jointly shaped by two policy parameters: the (shape of the) penalty function and the detection risk. With heterogeneous priors about the latter risk, for which we find several pieces of evidence, one cannot directly translate the bunching mass from these notches into a straightforward measure of behavioral elasticities. Hence, the evidence simply shows that (some) drivers rationally respond to the notched penalty scheme.

[29]Own computations based on data obtained from *Deutsches Statistisches Bundesamt* (*Fachserie 8/7, Verkehr*). Similar pre-post differences are obtained if one controls for month specific effects using de-trended monthly data.

In line with rational responses, our analysis also documents a significant change in the speed distribution after a reform of the penalty scheme. After the reform, which increased all fines for speeding 20km/h above the limit, we observe a 25% drop of drivers speeding in this range. At the same time, there is an increase in tickets with a speed slightly below the 20km/h cutoff.

From a normative perspective, these findings have several interesting implications. In principle, one might argue that a notched penalty scheme might be inferior as it is only a rough approximation to a Pigouvian correction mechanism. A 'true' Pigouvian mechanism would account for the fact that marginal externalities from speeding (accident risk, air and noise pollution, etc.) are continuously increasing in the speed level. When agents are imperfectly informed or boundedly rational, however, the simplicity of the stepwise scheme might increase awareness and contributes to the good knowledge of the penalty system – a point which is consistent with our survey evidence. The notched system could therefore dominate a more complex, Pigouvian penalty function. A further point that speaks in favor of the notched system directly follows from our bunching evidence: the fact that many drivers speed at similar speed levels tends to reduce the variance in speed. As suggested in Lave (1985), this could in turn contribute to a reduction in the accident risk. Ultimately, a credible and comprehensive welfare assessment which considers all these pros and cons requires exogenous variation between continuous and notched penalty schemes. Given the ubiquitousness of notches in law enforcement, seeking for such a quasi-experiment appears to be a promising direction for future research.

Appendix

3.A1. Proofs

Proof of Lemma 1. The result that the set of types of drivers that bunch at a certain notch x_i must be an interval $[\underline{\theta}_i, \overline{\theta}_i]$ is an immediate implication of the single crossing property: Let $\underline{\theta}_i := \inf\{\theta : x_i \in \arg\max EU(x;\theta)\}$ and $\overline{\theta}_i := \sup\{\theta : x_i \in \arg\max EU(x;\theta)\}$, and consider some $\theta \in [\underline{\theta}_i, \overline{\theta}_i]$. For all $x < x_i$, $[v(x_i, \theta) - pf(x_i)] - [v(x, \theta) - pf(x)] > [v(x_i, \underline{\theta}_i) - pf(x_i)] - [v(x, \underline{\theta}_i) - pf(x)] \geq 0$, and for all $x > x_i$, $[v(x, \theta) - pf(x)] - [v(x_i, \theta) - pf(x_i)] < [v(x, \overline{\theta}_i) - pf(x)] - [v(x_i, \overline{\theta}_i) - pf(x_i)] \leq 0$. Hence, all $\theta \in [\underline{\theta}_i, \overline{\theta}_i)$ strictly prefer x_i over any other speed level.

Let us now turn to characterizing the boundaries of that interval. As discussed in the main text, Figure 3.3 illustrates the case where $\underline{\theta}_i$ fulfills (4). Alternatively, however, x_i may be a corner solution for type $\underline{\theta}_i$, i.e. $x^*(\underline{\theta}_i) > x_i$. In this case, (5) claims that type $\underline{\theta}_i$ is indifferent between x_i and some other cutoff x_j, $j < i$. To prove this claim, suppose that type $\underline{\theta}_i$ strictly preferred x_i over any x_j, $j < i$. By continuity, there must be some type $\theta' < \underline{\theta}_i$ that also strictly prefers x_i over any x_j, $j < i$ and for which $x^*(\theta') > x_i$, which implies that θ' strictly prefers x_i over any $x < x_i$. Furthermore, by the single crossing property, $[v(x_i, \theta') - pf(x_i)] - [v(x, \theta') - pf(x)] > [v(x_i, \underline{\theta}_i) - pf(x_i)] - [v(x, \underline{\theta}_i) - pf(x)] \geq 0$ for all $x > x_i$. Hence, $\arg\max EU(x;\theta') = x_i$, a contradiction to the definition of $\underline{\theta}_i$.

As for the upper boundary $\overline{\theta}_i$, recall first that all $\theta < \overline{\theta}_i$ strictly prefer x_i over all $x > x_i$. Furthermore, for every $\theta > \overline{\theta}_i$, there exists an $x > x_i$ which this type θ strictly prefers over x_i. Hence, by continuity, there must be some $\hat{x} > x_i$ such that $v(\hat{x}, \overline{\theta}_i) - v(x_i, \overline{\theta}_i) = p(f_j - f_i)$, where f_j is the relevant fine for speed $\hat{x}$, i.e. $x_{j-1} < \hat{x} \leq x_j$ and $j > i$. As $\hat{x} \in \arg\max EU(x; \overline{\theta}_i)$, we have either $\hat{x} = x^*(\overline{\theta}_i) < x_j$, in which case (6) is satisfied, or $\hat{x} = x_j$, in which case (7) holds. ■

Proof of Lemma 2. For every θ and every $x, x' \leq x_\ell$ or $x, x' > x_\ell$, $EU(x;\theta) - EU(x';\theta)$ is not changed by the reform. However, the reform reduces $EU(x;\theta) - EU(x';\theta)$ for every $x' \leq x_\ell < x$ and every θ. Hence, if the reform induces any change in behavior, it can only be that drivers who drive at some speed $x > x_\ell$ before the reform, choose a speed $x \leq x_\ell$ after the reform. ■

Proof of Proposition 1. Let us first discuss the impacts of changing Δ_ℓ on the $\underline{\theta}_i$

and $\overline{\theta}_i$ of some cutoff x_i. Consider the effect on $\underline{\theta}_i$ and suppose that $\underline{\theta}_i$ is determined by condition (4). This equation is independent of f_ℓ, which implies that $\underline{\theta}_i$ remains unchanged. Suppose now that $\underline{\theta}_i$ is determined by condition (5). Taking the total differentials w.r.t. $\underline{\theta}_i$ and the difference $f_i - f_j$ yields

$$\frac{d\underline{\theta}_i}{d(f_i - f_j)} = \frac{p}{\frac{\partial v(x_i, \underline{\theta}_i)}{\partial \theta} - \frac{\partial v(x_j, \underline{\theta}_i)}{\partial \theta}} > 0 \tag{11}$$

due to the single-crossing property. An increase in Δ_ℓ has, *ceteris paribus*, an effect on the difference $f_i - f_j$ if and only if $j \leq \ell < i$. Furthermore, if it has any effect, then such an increase in Δ_ℓ will increase this difference $f_i - f_j$, thus increasing $\underline{\theta}_i$.

As for $\overline{\theta}_i$, suppose first that it is determined by (6). Taking the total differentials w.r.t. $\overline{\theta}_i$ and the difference $f_j - f_i$ yields

$$\frac{d\overline{\theta}_i}{d(f_j - f_i)} = \frac{p}{\frac{\partial v(x^*(\overline{\theta}_i), \overline{\theta}_i)}{\partial \theta} - \frac{\partial v(x_i, \overline{\theta}_i)}{\partial \theta}} > 0 \tag{12}$$

due to the single-crossing property. Suppose now that $\overline{\theta}_i$ is determined by (7). Taking the total differentials w.r.t. $\underline{\theta}_i$ and the difference $f_j - f_i$ yields

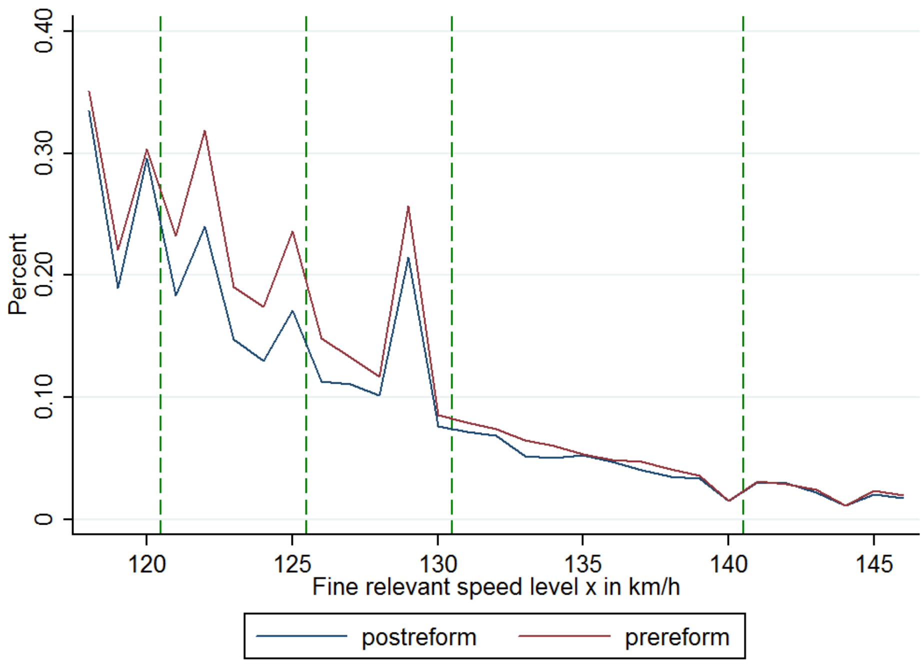
$$\frac{d\overline{\theta}_i}{d(f_j - f_i)} = \frac{p}{\frac{\partial v(x_j, \overline{\theta}_i)}{\partial \theta} - \frac{\partial v(x_i, \overline{\theta}_i)}{\partial \theta}} > 0 \tag{13}$$

again due to the single-crossing property. An increase in Δ_ℓ has, *ceteris paribus*, an effect on the difference $f_j - f_i$ if and only if $i \leq \ell < j$. Furthermore, if it has any effect, then such an increase in Δ_ℓ will increase this difference $f_\lambda - f_i$, thus increasing $\overline{\theta}_i$.

Using these results, we can complete the proofs:
Part (i): If $i > \ell$, an increase in Δ_ℓ has, *ceteris paribus*, no effect on $\overline{\theta}_i$ and may sometimes increase $\underline{\theta}_i$. Hence, such an increase in Δ_j weakly reduces Π_i.
Part (ii): If $i \leq \ell$, an increase in Δ_ℓ has, *ceteris paribus*, no effect on $\underline{\theta}_i$ and may sometimes increase $\overline{\theta}_i$. Hence, such an increase in Δ_j weakly increases Π_i. ■

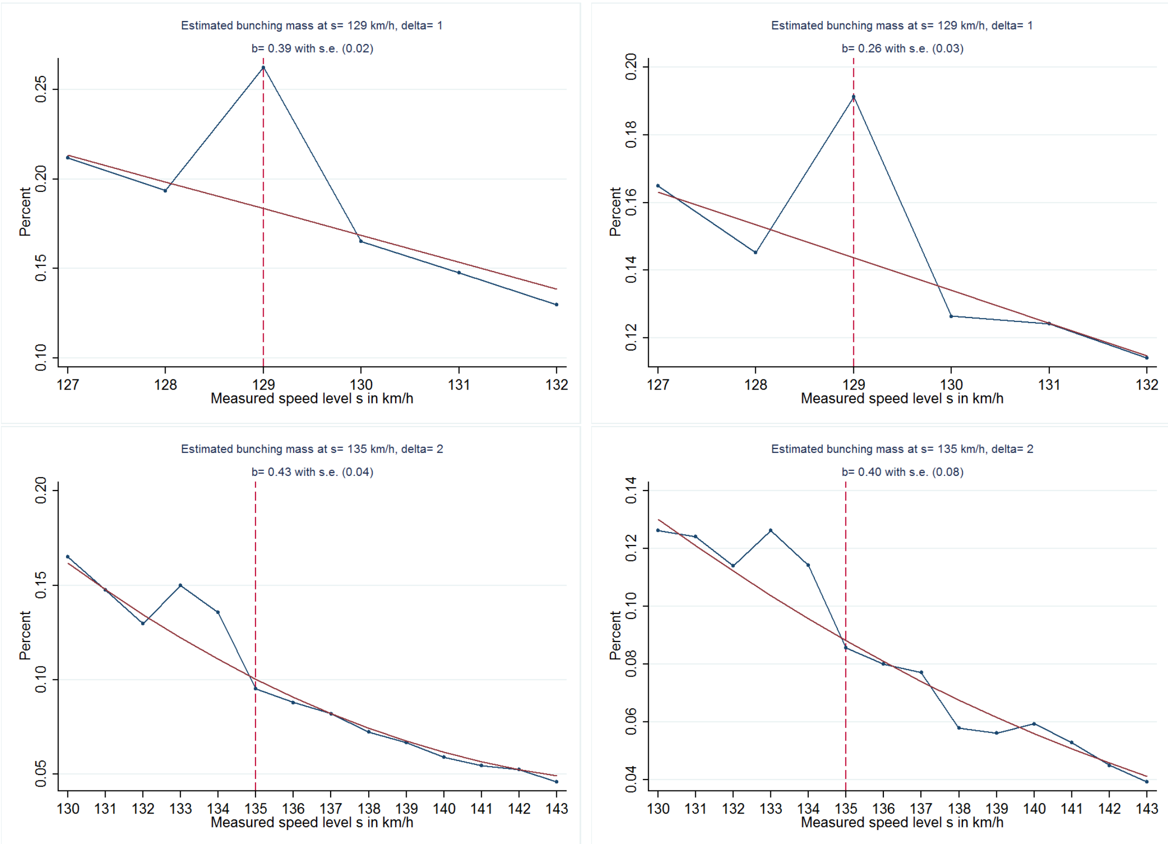
3.A2. Complementary Figures

Figure 3.6: Pre- and post-reform distribution of penalty-relevant speeding levels



Note: The figure illustrates the distribution of penalty-relevant speeding levels among speeding tickets from the pre- and post-reform period. The horizontal axis indicates the speed above the limit. The vertical axis indicates the percentage share of observations for each speed level. The green dashed vertical lines indicate the cutoffs at the respective speed levels. The number of observations for pre-reform [post-reform] period is 89,520 [34,356] in the displayed speed range.

Figure 3.7: Empirical and counterfactual distribution of speeding levels: Pre- and post-reform period



Note: The two panels on the left display the empirical and counterfactual distribution of measured speed for the *pre*-period. The panels on the right show the distributions for the *post*-reform period. The top panels compares the distribution around the cutoff $x_i = 125$ (corresponding to a measured speed $s = 129$km/h), the panels at the bottom consider the cutoff $x_i = 130$ (corresponding to $s = 135$km/h). The horizontal axes indicate the measured speed s. The vertical axes show the share of observations for each speed level in percent. The blue line captures the empirical distribution and the red curve shows the estimated counterfactual distribution. The counterfactual in the top [bottom] panels are estimated with a linear [quadratic] slope. The dashed vertical lines indicate the cutoffs in terms of measured speed s.

Supplementary Appendix

3.B. Online Survey: Complementary Material

3.B1. Main Survey Questions

1. **Consequences of Speeding**
In four sequential questions, participants are confronted with four different levels of speed, X_j, which are randomly and independently drawn from $X_1 \in \{19, 20, 21\}, X_2 \in \{24, 25, 26\}, X_3 \in \{29, 30, 31\}, X_4 \in \{34, 39\}$.

   Imagine you are notified about a speeding fine for a violation of the speed limit on a motorway. What is your estimate of the speeding fine for exceeding the speed limit by X_j kilometers per hour? (*Indicate the monetary fine in Euro; response in integer values.*)

2. **Penalty Increases**
At which values, if any at all, do the penalties for speeding increase? For example, select 19 km/h when you think the penalty for driving 19 km/h above the limit is greater than for a speed of 18 km/h above the limit. (*Please select only one of the response options per line.*)

19 km/h...	20 km/h...	21 km/h over limit	none of these values
24 km/h...	25 km/h...	26 km/h over limit	none of these values
29 km/h...	30 km/h...	31 km/h over limit	none of these values
39 km/h...	40 km/h...	41 km/h over limit	none of these values
49 km/h...	50 km/h...	51 km/h over limit	none of these values

3. **Avoiding a penalty**
In case of speeding, do you try to avoid higher penalties by staying under a certain threshold?

   Yes / More likely Yes / More likely No / No / I never speed

4. **Relevant Speed for determining Penalty (1)**
Do you know the official deduction rule (*Toleranzabzug Regel*) for computing the speed level which is then relevant for determining the penalties?

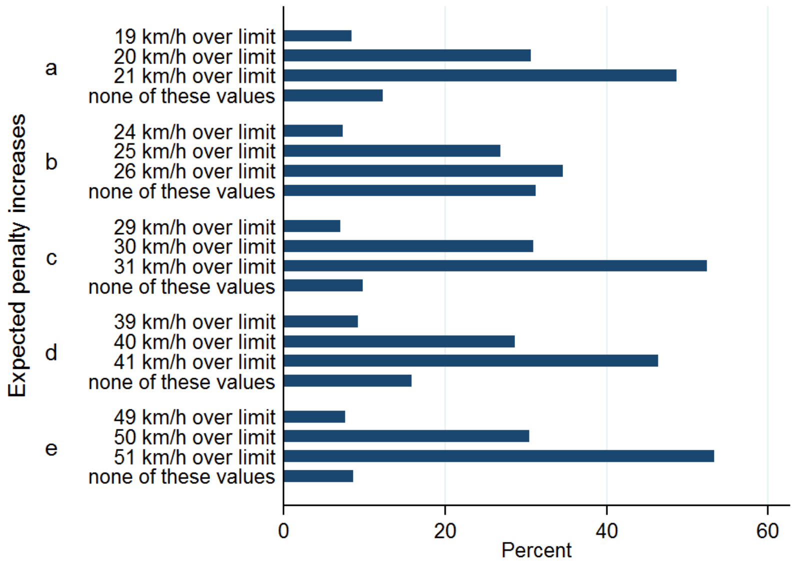
Yes / Not exactly / No

5. **Relevant Speed for determining Penalty (2)**
   Subjects who answered the previous question with 'No' skipped this one.
   By how many percent does the official deduction rule subtract from the measured speed?

   By 0% / 1% / ... / 9% / 10% (11 options)

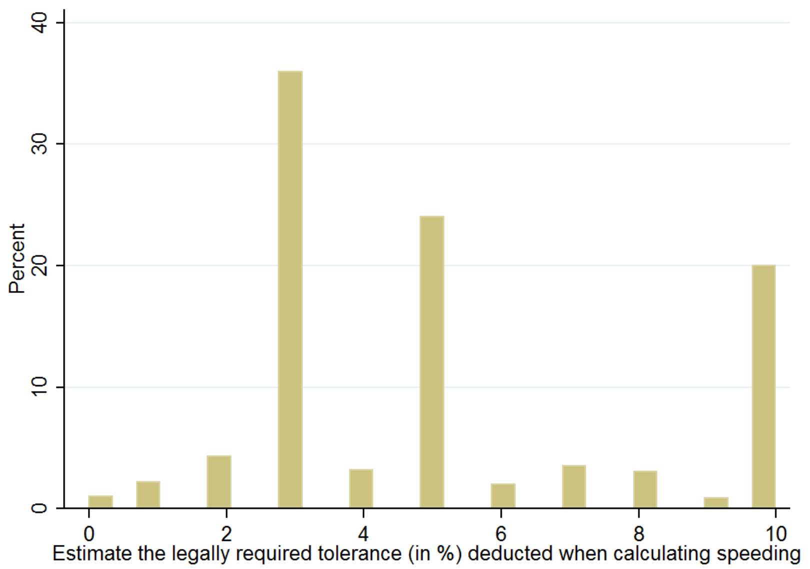
3.B2. Complementary Survey Evidence

Figure 3.8: Expectations regarding the cutoff points



Notes: For each of the five cutoffs (21, 26, 31, 41 and 51km/h above the speed limit), the bar graphs represent the fraction of respondents who expect an increase in the penalties at one (or none) of the indicated speed levels. (See survey question 2 from Online Survey, above.)

Figure 3.9: Expected tolerance rule deductions (in percent)



Notes: The bar graph represents the fraction of respondents who expect the tolerance rule to deduct 0, 1, ..., 10 percent of the measured speed. (See survey question 5 from Online Appendix, above.)

References

Achenbach, Thomas M., Andreas Becker, Manfred Döpfner, Einar Heiervang, Veit Roessner, Hans-Christoph Steinhausen, and Aribert Rothenberger (2008), 'Multicultural assessment of child and adolescent psychopathology with ASEBA and SDQ instruments: research finding, applications, and future directions', *Journal of Child Psychology and Psychiatry*, Vol. 49(3), p. 251-275.

Adda, Jérôme, James Banks, and Hans-Martin Von Gaudecker (2009), 'The impact of income shocks on health: Evidence from cohort data', *Journal of the European Economic Association*, Vol. 7(6), p. 1361-1399.

Adda, Jérôme, Anders Björklund, and Helena Holmlund (2011), 'The role of mothers and fathers in providing skills: Evidence from parental deaths', IZA Discussion Paper, 5425.

Anderson, D. Mark (2014), 'In school and out of trouble? The minimum dropout age and juvenile crime', *Review of Economics and Statistics*, Vol. 96(2), p. 318-331.

Anderson, D. Mark, and Mary Beth Walker (2015), 'Does shortening the school week impact student performance? Evidence from the four-day school week', *Education Finance and Policy*, Vol. 10(3), p. 314-349.

Andrews, Rodney J., and Trevon D. Logan (2010), 'Family health, children's own health and test score gaps', *American Economic Review, Paper and Proceedings*, Vol. 100, p. 195-199.

Åslund, Olof, Hans Grönqvist, Caroline Hall, and Jonas Vlachos (2015), 'Education and criminal behavior: Insights from an expansion of upper secondary school', IZA Discussion Paper, 9374.

Balsa, Ana I. (2008), 'Parental problem-drinking and adult children's labor market outcomes', *Journal of Human Resources*, Vol. 43(2), p. 454-486.

Bastani, Spencer, and Håkan Selin (2014), 'Bunching and non-bunching at kink points of the Swedish tax schedule', *Journal of Public Economics*, Vol. 109, p. 36-49.

Baumgärtner, Theo, and Johannes Kestler (2014), 'Suchtmittelgebrauch, Computerspielverhalten, Internetnutzung und Glücksspielerfahrungen von Jugendlichen in Hamburg und drei Modellregionen in Deutschland. Deskriptive Ergebnisse der SCHULBUS-regional-Studie 2012', HLS/BfS-Berichte, SB 14-B1.

Becker, Andreas, Hans-Christoph Steinhausen, Gisli Baldursson, Sören Dalsgaard, Maria J. Lorenzo, Stephen J. Ralston, Manfred Döpfner, and Aribert Rothenberger (2006), 'Psychopathological screening of children with ADHD: Strengths and Difficulties Questionnaire in a pan-European study', *European Child & Adolescence Psychiatry*, Vol. 15(1), p. 56-62.

Becker, Gary S. (1968), 'Crime and punishment: An economic approach', *Journal of Political Economy*, Vol. 76(2), p. 169-217.

Berger, Eva M., Frauke H. Peter, and C. Katharina Spieß (2010), 'Wie hängen familiäre Veränderungen und das mütterliche Wohlbefinden mit der frühkindlichen Entwicklung zusammen?', *Quarterly Journal of Economic Research*, Vol. 79(3), p. 27-44.

Berger, Eva M., and C. Katharina Spieß (2011), 'Maternal life satisfaction and child outcomes: Are they related?', *Journal of Economic Psychology*, Vol. 32(1), p. 142-158.

Berthelon, Matias E., and Diana I. Kruger (2011), 'Risky behavior among youth: Incapacitation effects of school on adolescent motherhood and crime in Chile', *Journal of Public Economics*, Vol. 95(1), p. 41-53.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan (2004), 'How much should we trust Differences-in-Differences estimates?', *Quarterly Journal of Economics*, Vol. 119(1), p. 249-275.

Blinder, Alan, and Harvey S. Rosen (1985), 'Notches', *American Economic Review*, Vol. 75, p. 736-747.

Blomeyer, Dorothea, Katja Coneus, Manfred Laucht, and Friedhelm Pfeiffer (2009), 'Initial risk matrix, home resources, ability development and children's achievement', *Journal of the European Economic Association*, Vol. 7(2-3), p. 638-648.

Bourgeon, Jean-Marc, and Pierre Picard (2007), 'Point-record driving licence and road safety: An economic approach', *Journal of Public Economics*, Vol. 91(1-2), p. 235-258.

Brilli, Ylenia, and Marco Tonello (2015), 'Rethinking the crime reducing effect of education? Mechanisms and evidence from regional divides', Bank of Italy, Economic Research and International Relations Area, 1008.

Büttner, Bettina, and Stephan L. Thomsen (2015), 'Are we spending too many years in school? Causal evidence of the impact of shortening secondary school duration', *German Economic Review*, Vol. 16(1), p. 65-86.

Cameron, A. Colin, and Douglas L. Miller (2015), 'A practitioner's guide to cluster-robust inference', *Journal of Human Resources*, Vol. 50(2), p. 317-372.

Case, Anne, and Christina Paxson (2005), ‘Sex differences in morbidity and mortality’, *Demography*, Vol. 42(2), p. 189-214.

Castillo-Manzano, José I., Mercedes Castro-Nuño, and Diego J. Pedregal (2010), ‘An econometric analysis of the effects of the penalty points system driver’s license in Spain’, *Accident Analysis & Prevention*, Vol. 42(4), p. 1310-1319.

Chetty, Raj (2012), ‘Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply’, *Econometrica*, Vol. 80, p. 969-1018.

Chetty, Raj, John N. Friedman, Tore Olsen, and Luigi Pistaferri (2011), ‘Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from Danish tax records’, *Quarterly Journal of Economics*, Vol. 126, p. 749-804.

Chetty, Raj, John N. Friedman, and Emmanuel Saez (2013), ‘Using differences in knowledge across neighborhoods to uncover the impacts of the EITC on earnings’, *American Economic Review*, Vol. 103(7), p. 2683-2721.

Chevalier, Arnaud, and Olivier Marie (2013), ‘Economic uncertainty, parental selection, and the criminal activity of the ‘children of the wall’ ’, CESifo Working Paper, 4462.

Chevalier, Arnaud, and Olivier Marie (2016), ‘Economic uncertainty, parental selection, and children’s educational outcomes’, *Journal of Political Economy*, forthcoming.

Ciccia, Rossella, and Mieke Verloo (2012), ‘Parental leave regulations and the persistence of the male breadwinner model: Using fuzzy-set ideal type analysis to assess gender equality in an enlarged Europe’, *Journal of European Social Policy*, Vol. 22(5), p. 507-528.

Coneus, Katja, Andrea M. Mühlenweg, and Holger Stichnoth (2014), ‘Orphans at risk in sub-Saharan Africa: Evidence on educational and health outcomes’, *Review of Economics of the Household*, Vol. 12(4), p. 641-662.

Congdon, William J., Sendhil Mullainathan, and Jeffrey R. Kling (2011), *Policy and choice: Public finance through the lens of behavioral economics*, Washington, DC: Brookings Institution Press.

Cunha, Flavio, James J. Heckman, Lance Lochner, and Dimitriy V. Masterov (2006), 'Interpreting the evidence on life cycle skill formation', in A. Hanushek and F. Welch (Eds.), *Handbook of the Economics of Education* (p. 697-812), Amsterdam: North-Holland.

Cunha, Flavio, and James J. Heckman (2008), 'Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation', *Journal of Human Resources*, Vol. 43(4), p. 738-782.

Currie, Janet, and Brigitte C. Madrian (1999), 'Health, health insurance and the labor market', in O. Ashenfelter and R. Layard (Eds.), *Handbook of Labor Economics* (p. 3309-3416), Amsterdam: North-Holland.

Currie, Janet, and Douglas Almond (2011), 'Human capital development before age five', in O. Ashenfelter and R. Layard (Eds.), *Handbook of Labor Economics* (p. 1315-1486), Amsterdam: Elsevier Science B.V..

Dahmann, Sarah, and Silke Anger (2014), 'The impact of education on personality: Evidence from a German high school reform', IZA Discussion Paper, 8139.

DeAngelo, Gregory J., and Benjamin Hansen (2014), 'Life and death in the fast lane: Police enforcement and traffic fatalities', *American Economic Journal: Economic Policy*, Vol. 6(2), p. 231-257.

De Bartolome, Charles (1995), 'Which tax rate do people use: Average or marginal?', *Journal of Public Economics*, Vol. 56(1), p. 79-96.

Deming, David James (2011), 'Better schools, less crime?', *The Quarterly Journal of Economics*, Vol. 126, p. 2063-2115.

De Paola, Maria, Vincenzo Scoppa, and Mariatiziana Falcone (2013), 'The deterrent effects of penalty point system in driving licenses: A regression discontinuity approach', *Empirical Economics*, Vol. 45(2), p. 965-985.

Dooley, Martin, and Jennifer Stewart (2007), 'Family income, parenting styles and child behavioural-emotional outcomes', *Health Economics*, Vol. 16(2), p. 145-162.

Drogenbeauftragte der Bundesregierung (2013), 'Drogen-und Suchtbericht', Berlin, Germany: Federal Ministry of Health.

Dusek, Libor, and Christian Traxler (2016), 'Experience with punishment and specific deterrence: Evidence from speeding tickets', Mimeo.

Dustmann, Christian (2004), 'Parental background, secondary school track choice, and wages', *Oxford Economic Papers*, Vol. 56(2), p. 209-230.

Ehrlich, Isaac (1975), 'On the relation between education and crime', in F. Thomas Juster (Ed.), *Education, income, and human behavior* (p. 313-337), New York: McGraw-Hill.

Ermisch, John, Frauke H. Peter, and C. Katharina Spieß (2012), 'Early childhood outcomes and family structure', in J. Ermisch, M. Jäntti, and T. Smeeding (Eds.), *From Parents to Children: The Intergenerational Transmission of Advantage* (p. 120-139), New York: Russell Sage Foundation.

Farahati, Farah, Dave E. Marcotte, and Virginia Wilcox-Gök (2003), 'The effects of parents' psychiatric disorders on children's high school dropout', *Economics of Education Review*, Vol. 22(2), p. 167-178.

Feldman, Naomi E., Peter Katuscak, and Laura Kawano (2015), 'Taxpayer confusion: evidence from the child tax', *American Economic Review*, forthcoming.

Frank, Richard G., and Ellen Meara (2009), 'The effect of maternal depression and substance abuse on child human capital development', NBER Working Paper Series, 15314.

Gillitzer, Christian, Henrik Kleven, and Joel Slemrod (2016), 'A characteristics approach to optimal taxation: line drawing and tax-driven product innovation', *Scandinavian Journal of Economics*, forthcoming.

Goodman, Robert (1997), 'The Strengths and Difficulties Questionnaire: A research note', *Journal of Child Psychology and Psychiatry*, Vol. 38(5), p. 581-586.

Graves, Philip, Dwight Lee, and Robert Sexton (1993), 'Speed variance, enforcement, and the optimal speed limit', *Economics Letters*, Vol. 42, p. 237-243.

Hagan, Ronald, Andrew M. Jones, and Nigel Rice (2009), 'Health and retirement in Europe', *International Journal of Environmental Research and Public Health*, Vol. 6(10), p. 2676-2695.

Hansen, Benjamin (2015), 'Punishment and deterrence: Evidence from drunk driving', *American Economic Review*, Vol. 105(4), p. 1581-1617.

Heymann, Jody, Hye Jin Rho, John Schmitt, and Alison Earle (2009), 'Contagion nation: a comparison of paid sick day policies in 22 countries', Center for Economic and Policy Research (CEPR), 2009-19.

Huebener, Mathias, and Jan Marcus (2015), 'Moving up a gear: The impact of compressing instructional time into fewer years of schooling', DIW Discussion Paper, 1450.

Jäncke, Lutz, Beatrice Brunner, and Michaela Esslen (2008), 'Brain activation during fast driving in a driving simulator: the role of the lateral prefrontal cortex', *Cognitive Neuroscience and Neuropsychology*, Vol. 19(11), p. 1127-1130.

Johnson, Eric, and C. Lockwood Reynolds (2013), 'The effect of household hospitalizations on the educational attainment of youth', *Economics of Education Review*, Vol. 37, p. 165-182.

Jondrow, James, Marianne Bowes, and Robert Levy (1983), 'The optimal speed limit', *Economic Inquiry*, Vol. 21, p. 325-336.

Jones, Andrew M., Nigel Rice, and Jennifer Roberts (2010), 'Sick of work or too sick to work? Evidence on self-reported health shocks and early retirement from the BHPS', *Economic Modelling*, Vol. 27(4), p. 866-880.

Kim-Cohen, Julia, Terrie E. Moffitt, Alan Taylor, Susan J. Pawlby, and Avshalom Caspi (2005), 'Maternal depression and children's antisocial behavior: nature and nurture effects', *Archives of General Psychiatry*, Vol. 62(2), p. 173-181.

Kleven, Henrik (2016), 'Bunching', *Annual Review of Economics*, forthcoming.

Kleven, Henrik, and Mazhar Waseem (2013), 'Using Notches to uncover optimization frictions and structural elasticities: Theory and evidence from Pakistan', *Quarterly Journal of Economics*, Vol. 128(2), p. 669-723.

Kline, Patrick (2012), 'The impact of juvenile curfew laws on arrests of youth and adults', *American Law and Economics Review*, Vol. 14(1), p. 44-67.

Kühn, Svenja M., Isabell van Ackeren, Gabriele Bellenberg, Christian Reintjes, and Grit im Brahm (2013),'Wie viele Schuljahre bis zum Abitur? Eine multiperspektivische Standortbestimmung im Kontext der aktuellen Schulzeitdebatte', *Zeitschrift für Erziehungswissenschaften*, Vol. 16, p. 115-136.

Lave, Charles A. (1985), 'Speeding, coordination, and the 55 MPH limit', *American Economic Review*, Vol. 75, p. 1159-1164.

Lefgren, Lars, and Brian A. Jacob (2003), 'Are idle hands the devil's workshop? Incapacitation, concentration, and juvenile crime', *American Economic Review*, Vol. 93(5), p. 1560-1577.

Liebman, Jeffrey B., and Richard J. Zeckhauser (2004), 'Schmeduling', Working Paper, Harvard University.

Lochner, Lance, and Enrico Moretti (2004), 'The effect of education on crime: Evidence from prison inmates, arrests, and self-reports', *American Economic Review*, Vol. 94(1), p. 155-189.

Luallen, Jeremy (2006), 'School's out... forever: A study of juvenile crime, at-risk youths and teacher strikes', *Journal of Urban Economics*, Vol. 59(1), p. 75-103.

Machin, Stephen, Olivier Marie, and Sunčica Vujić (2011), 'The crime reducing effect of education', *The Economic Journal*, Vol. 121(552), p. 463-484.

Meyer, Tobias, Stephan L. Thomsen, and Heidrun Schneider (2015), 'New evidence on the effects of the shortened school duration in the German states: An evaluation of post-secondary education decisions', IZA Discussion Papers, 9507.

Milde-Busch, Astrid, A. Blaschek, I. Borggräfe,R. von Kries, A. Straube and F. Heinen (2010), 'Besteht ein Zusammenhang zwischen der verkürzten Gymnasialzeit und Kopfschmerzen und gesundheitlichen Belastungen bei Schülern im Jugendalter?', *Klinische Pädiatrie*, Vol. 222(04), p. 1-6.

Montag, Josef (2014), 'A radical change in traffic law: Effects on fatalities in the Czech Republic', *Journal of Public Health*, Vol. 36, p. 539-545.

Morefield, Brant (2010), 'Parental health events and children's skill development', Working Papers 10-11, University of North Carolina at Greensboro, Department of Economics.

OECD (2012), 'The distribution of working hours among adults in couple families by age of the youngest child and number of children', OECD family database, LMF.2.2, OECD Social Policy Division, Accessed 19 January 2016, www.oecd.org/els/family/LMF2_2 _Usual_working_hours_of_couple_parents_Sep2013.pdf.

Peden, Margie, Richard Scureld, David Sleet, Dinesh Mohan, Andnan A. Hyder, Eva Jarawan, and Colin Mathers (2004), 'World report on road traffic injury prevention', World Health Organization, Geneva.

Petterson, Stephen M., and Alison Burke Albers (2001), 'Effects of poverty and maternal depression on early child development', *Child Development*, Vol. 72(6), p. 1794-1813.

Podor, Melinda, and Timothy J. Halliday (2012), 'Health status and the allocation of time', *Health Economics*, Vol. 21(5), p. 514-527.

Propper, Carol, John Rigg, and Simon Burgess (2007), 'Child health: evidence on the roles of family income and maternal mental health from a UK birth cohort', *Health Economics*, Vol. 16(11), p. 1245-1269.

Rasmusen, Eric B. (1995), 'How optimal penalties change with the amount of harm', *International Review of Law and Economics*, Vol. 15, p. 101-108.

Ribar, David (2004), 'What do social scientists know about the benefits of marriage? A review of quantitative methodologies', IZA Discussion Paper, 998.

Riphahn, Regina T. (1999), 'Income and employment effects of health shocks A test for the German welfare state', *Journal of Population Economics*, Vol. 12(3), p. 363-389.

Ruhm, Christopher J. (2004), 'Parental employment and child cognitive development', *Journal of Human Resources*, Vol. 39(1), p. 155-192.

Saez, Emmanuel (2010), 'Do taxpayers bunch at kink points?', *American Economic Journal: Economic Policy*, Vol. 2, p. 180-212.

Sallee, James M., and Joel Slemrod (2012), 'Car notches: Strategic automaker responses to fuel economy policy', *Journal of Public Economics*, Vol. 96, p. 981-999.

Schmiade, Nicole, C. Katharina Spieß, and Wolfgang Tietze (2008), 'Zur Erhebung des adaptiven Verhaltens von zwei- und dreijährigen Kindern im Sozio-oekonomischen Panel (SOEP)', SOEPpapers on Multidisciplinary Panel Data Research, 116.

Schneider, Henning (2008), 'Natürliche Geburt oder "Wunsch-Sectio"?', *Der Gynäkologe*, Vol. 41, p. 36-41.

Senne, Jean-Noël (2014), 'Death and schooling decisions over the short and long run in rural Madagascar', *Journal of Population Economics*, Vol. 27(2), p. 497-528.

Slemrod, Joel (2013), 'Buenas notches: Lines and notches in tax system design', *eJournal of Tax Research*, Vol. 11, p. 259-283.

Smith, James P. (1999), 'Healthy bodies and thick wallets: The dual relation between health and economic status', *The Journal of Economic Perspectives*, Vol. 13(2), p. 145-166.

Smith, James P. (2004), 'Unraveling the SES-Health Connection', *Population and development review, Supplement: Aging, health, and public policy*, Vol. 30, p. 108-132.

Smith, James P. (2005), 'Consequences and predictors of new health events', in D.A. Wise (Ed.), *Analysis in the Economics of Aging* (p. 213-237), Chicago: University of Chicago Press.

Sun, Ang, and Yang Yao (2010), 'Health shocks and children's school attainments in rural China', *Economics of Education Review*, Vol. 29(3), p. 375-382.

Todd, Petra E., and Kenneth I. Wolpin (2007), 'The production of cognitive achievement in children: Home, school, and racial test score gaps', *Journal of Human Capital*, Vol. 1(1), p. 91-136.

Tonks, James, Alan Slater, Ian Frampton, Sarah E. Wall, Phil Yates, and W. Huw Williams (2009), 'The development of emotion and empathy skills after childhood brain injury', *Developmental Medicine & Child Neurology*, Vol. 51(1), p. 8-16.

UNESCO (2006), 'International standard classification of education, ISCED 1997', May 2006 Re-edition, UNESCO-UIS.

van Bentham, Arthur (2015), 'What is the optimal speed limit on freeways?', *Journal of Public Economics*, Vol. 124, p. 44-62.

Violato, Mara, Stavros Petrou, Ron Gray, and Maggie Redshaw (2011), 'Family income and child cognitive and behavioural development in the United Kingdom: does money matter?', *Health Economics*, Vol. 20(10), p. 1201-1225.

Wagner, Gert G., Joachim R. Frick, and Jürgen Schupp (2007), 'The German socio-economic panel study (SOEP) - Scope, evolution and enhancements', *Schmollers Jahrbuch*, Vol. 127(1), p. 39-169.

Wu, Stephen (2003), 'The Effects of health events on the economic status of married couples', *Journal of Human Resources*, Vol. 38(1), p. 219-230.

Curriculum Vitae

September 2013 – January 2016	Ph.D. program in governance Hertie School of Governance, Berlin
May 2011	Diploma in economics Freie Universität Berlin
October 2005 - May 2011	Studies in economics Freie Universität Berlin
August 2002 - March 2005	High School (Abitur) Privates Gymnasium der Zisterzienser Abtei Marienstatt
12 August 1985	born in Hachenburg

Publication

The 1st chapter of this dissertation has been published under the following reference:

Mühlenweg, A.M., Westermaier, F.G. and Morefield, B., (2015) 'Parental health and child behavior: evidence from parental health shocks', *Review of Economics of the Household*, p. 1-22.