Education, Health Risks, and Economic Outcomes –

Essays Based on Micro Data

INAUGURAL-DISSERTATION

zur Erlangung des akademischen Grades

eines Doktors der Wirtschaftswissenschaft

(doctor rerum politicarum)

des Fachbereichs Wirtschaftswissenschaft

der Freien Universität Berlin

vorgelegt von

Diplom-Volkswirt Daniel Kemptner

Berlin, 2013

Erstgutachter:

Prof. Peter Haan, Freie Universität Berlin

Zweitgutachter:

Prof. Frank Fossen, Freie Universität Berlin

Tag der Disputation:

27.06.2013

Erklärung über Zusammenarbeit mit Koautoren und Vorveröffentlichungen

Kapitel 1:

Jürges, H., D. Kemptner und S. Reinhold (2011). Changes in compulsory schooling and the causal effect of education on health: Evidence from Germany. *Journal of Health Economics*, 30(2), 340 - 354.

Kapitel 2:

Kemptner, D. und J. Marcus (2013). Spillover effects of maternal education on child's health and health behavior. *Review of Economics of the Household*, 11(1), 29 - 52.

Kapitel 3:

Kemptner, D., P. Haan und A. Uhlendorff (2012). Bayesian procedures as a numerical tool for the estimation of dynamic discrete choice models. *DIW Discussion Papers*, 1210.

Kapitel 4:

unveröffentlicht

Education, Health Risks, and Economic Outcomes –

Essays Based on Micro Data

Thesis

Acknowledgements

The Graduate Center of DIW Berlin gave me the great opportunity to write my doctoral thesis at a research institute. Furthermore, it provided me with academic training, financial support, and a wonderful academic community of young researchers. Since the second year of my doctoral studies, I worked in the Public Economics department of DIW Berlin. I benefited from the nice atmosphere and the helpfulness of all my dear colleagues in the department. I am also grateful for the financial support of Thyssen foundation.

Most of all, I like to thank my supervisor, Peter Haan, for his guidance, valuable comments and continuous support. I also like to thank my second supervisor, Frank Fossen. Furthermore, I am grateful to my coauthors Steffen Reinhold, Hendrik Jürges, Jan Marcus, Peter Haan, and Arne Uhlendorf for a productive and enjoyable collaboration. I like to thank Viktor Steiner for his support at the beginning of my doctoral studies.

I dedicate this thesis to my parents, Udo and Gabriele, and my brother, Christoph, who always backed me up.

Contents

General Introduction

This thesis comprises four chapters and investigates the relationship between individuals' education, health risks, and economic outcomes in Germany. The focus is set on i) a robust identification of the causal effect of education on health risks and ii) the dynamic relationship between health and economic outcomes over the life-cycle. Health and education are two main components of human capital and their positive association is a robust finding in the empirical literature (see David and Lleras-Muney (2012) or Grossman (2006) for overviews). A good understanding of their relationship is highly policy-relevant. Among the factors of socio-economic status (education, income, and occupation), education seems to be the one that - in the long run - is most amenable to public policy interventions. If education improves both health and economic outcomes, giving individuals access to more and better education may be more efficient than redistributing income or expanding public health care expenditures. As noted by Deaton (2002), it is important to frame social policy in terms of health and wealth simultaneously. Since health shocks affect the dynamics of labor market participation, early retirement, and wealth accumulation, a structural econometric analysis of health and economic outcomes should take into account the relevant processes and the individuals' expectations about the consequences of their decisions.

The main contributions of this thesis to the economic literature are twofold: i) Robust evidence for the causal link between education and health risks is provided and ii) the implications of health risks for individuals' economic outcomes are quantified from a life-cycle perspective. The evidence allows drawing conclusions for public policy. Education is a means to improve not only income opportunities but also health outcomes. There is even a causal link between maternal education and children's health behavior. This suggests that education policy is also health policy. There are non-monetary gains of education that should be taken into account when designing an optimal education policy.

Simulations on the basis of a structural life-cycle model show that there are substantial health-related consumption risks that are uninsured by the German social security system. Moreover, there is a concern of health-related old age poverty. These findings suggest that there may be need for either additional private disability insurance or better public insurance through the early retirement option of the statutory pension insurance scheme. When considering a public intervention, the introduction of means-tested minimum pension benefits appears to be a practicable approach in order to protect individuals from the risk of old age poverty.

The research of the first two chapters resorts to reduced form approaches in order to robustly identify the causal effects of education on health outcomes. These effects are examined with respect to both an individual's own health (chapter 1) and in the intergenerational context (chapter 2). Despite many convincing arguments why there should be a causal effect of education on health, concerns about an interpretation of correlations as causal relationships remain. Most importantly, unobserved heterogeneity could drive both educational attainment at younger ages and health outcomes at older ages. Furthermore, there may be reversed causality if an individual's health affects educational outcomes. This involves serious identification problems. Therefore, the research that is presented in the first two chapters resorts to instrumental variable methods in order to deal with the potential endogeneity of education when estimating the effects. These methods generally aim at exploiting an exogenous variation in the endogenous variable (here: education) to achieve identification. Various sensitivity checks and an investigation of potential channels of the estimated effects complement the analysis.

In chapters 3 and 4, structural econometric models are estimated to explain economic outcomes on the basis of behavioral economic models. While this requires more (other) assumptions than reduced form approaches, the estimation of structural parameters sheds light on the underlying mechanisms and allows the - ex ante - simulation of scenarios without the need for quasi-experimental evidence. Estimation of complex structural models may be computationally burdensome and considerable numerical problems with classical estimation procedures may arise (e.g. non-convergence or convergence at a local maximum of the optimization algorithm). Therefore, chapter 3 discusses the numerical advantage of using Bayesian estimation procedures for the estimation of dynamic discrete choice models. This becomes relevant for the research in chapter 4, where an extension of the Expectation Maximization algorithm is applied to facilitate the estimation of a complex dynamic model of health risks, labor market participation, early retirement, and wealth accumulation. Similarly to Bayesian estimation procedures, the Expectation Maximization algorithm realizes numerical advantages by exploiting information on the posterior distribution of the model's parameters. Subsequently, the model is used to investigate the health-related risks of consumption and old age poverty.

Chapter 1 contributes to the growing literature on identifying the causal link between education and health and health-related behavior. The study estimates the causal effect of schooling on health exploiting variations in compulsory schooling laws in West Germany. We follow the approach of Pischke and von Wachter (2005, 2008) who have used changes in years of compulsory schooling to estimate the causal effect of education on earnings. The authors found zero returns to schooling on wages. However, when estimating a regression of body mass index on the change in compulsory schooling laws they found a significant effect. Our objective is to take this line of research further and assess the non-monetary returns of education by considering other health outcomes and also important dimensions of health-related behavior such as smoking. Using data from several waves of the German Microcensus, we exploit changes in years of compulsory schooling in West German federal states that took effect between 1949 and 1969 to estimate the causal effect of years in school on long-term illness, work disability, body mass index (and overweight/obesity) and current and former smoking measured in 1989 to 2003. We address possible concerns about the validity of our results that stem from a potential migration bias, from unobserved state-specific trends, from potential non-monotonicity of the relationship between the change in compulsory schooling laws and years of schooling, and from the introduction of so-called short school years. Besides, we examine health inputs (smoking, overweight, obesity) and occupation (manual vs. non-manual) as mediating factors of the estimated effects.

Chapter 2 investigates the effects of maternal education on child's health and health behavior in Germany. A quantification of such intergenerational links is not only relevant regarding optimal investments into education, but also relates to social mobility. We add to the literature i) by applying an instrumental variables strategy that works for the sample size of common household panels and that has not yet been used in the intergenerational context, ii) by considering a wide range of outcomes for both newborns and adolescents, iii) and by investigating possible channels of the estimated effects. For this purpose, we draw on a rich household survey, the German Socio-Economic Panel Study, containing detailed information about three generations. We instrument maternal education by the number of her siblings while conditioning on characteristics of her parents, the child's grandparents. We argue that, given the grandparents' characteristics, the number of the mother's siblings generates variation in maternal years of education that is exogenous regarding her child's health and health behavior. If grandparents are constrained in borrowing against the mother's future earnings, the number of her siblings affects household resources available for her educational investments. We demonstrate the robustness of our instrumental variable estimates in a variety of sensitivity checks regarding control variables, sample characteristics, functional form, and assumptions of the error term. Furthermore, we examine mother's health behavior, assortative mating, household income, and child's schooling track as potential channels of the estimated effects.

Chapter 3 provides evidence for the advantage of using Bayesian estimation procedures instead of classical maximum likelihood estimation for the estimation of dynamic discrete choice models. Accounting for unobserved heterogeneity is important when estimating non-linear models. Numerous studies document that discrete choice models without unobserved heterogeneity require either very strong and often implausible assumptions or lead to biased estimates of central parameters. In particular, this is true for dynamic models which analyze the role of state dependence in the behavior of agents because it is necessary to disentangle true state dependence from individual specific effects. These models usually require a general specification of unobserved preference heterogeneity and often relatively complex estimation routines need to be applied. After a general discussion of both classical and Bayesian estimation procedures for the estimation of discrete choice models, we consider an application of a dynamic discrete choice model of female labor supply with three distinct states and estimate dynamic mixed logit models. Our analysis is based on longitudinal data from the German Socio-Economic Panel. The findings motivate the application of an extension of the Expectation Maximization algorithm in chapter 4 because - similarly to Bayesian estimation procedures - this algorithm exploits information on the posterior distribution of the model's parameters.

Chapter 4 proposes a rich life-cycle to investigate the health-related risks of consumption and old age poverty. The study adds to the literature in several ways. First, it demonstrates the good performance of an extension of the

Expectation-Maximization algorithm by estimating a complex dynamic model of health risks, labor market participation, early retirement, and wealth accumulation. The extension of the Expectation-Maximization algorithm is used to obtain good starting values for a subsequent full information maximum likelihood estimation. I rely on the framework of a dynamic programming discrete choice model that is estimated using data from the German Socio-Economic Panel Study. Second, I simulate scenarios where health shocks do or do not occur at different ages during the life-cycle for individuals with differing endowments. A comparison of consumption paths and net present values of expected lifetime consumption between the scenarios sheds light on health-related consumption and poverty risks that are uninsured by the German social security system. Third, I simulate the introduction of minimum pension benefits that protect individuals from the risk of old age poverty. Since this raises a concern regarding an increase in abuse of the early retirement option and a decrease in average pension age, I examine whether a means test may reduce the potential increase in abuse. The implications of the analysis may also apply to other countries with similar institutions.

Chapter 1

Changes in compulsory schooling and the causal effect of education on health: Evidence from Germany†

1.1 Introduction

One of the most robust findings in both the economic and medical literature is the positive association between education and health (e.g. see the survey articles by Cutler and Lleras-Muney, 2008; Grossman, 2006). This relationship can be found in many countries, at different education levels and for various indicators of health. In fact, the relationship between education or socio-economic status in a wider sense and health is so ubiquitous that is often simply referred to as "the" gradient (Deaton (2003)). However, the association between education and health behavior does not necessarily reflect a causal effect of education on health and there is now a lively debate, especially in the economics literature, whether this association (or how much of it) is causal.

This discussion is highly policy-relevant because, among the components of socio-economic status (education, income, and occupation), education seems to be the one that - in the long run - appears to be most amenable to public policy interventions. Moreover, as noted by Deaton (2002), it is important to

†This chapter is based on joint work with Steffen Reinhold and Hendrik Jürges, see Kemptner et al. (2011).

frame social policy in terms of health and wealth simultaneously. Improving one at the expense of the other involves a difficult and probably unnecessary trade-off. One possible policy instrument for improving health and wealth simultaneously is education. If education improves both wealth and health, giving people access to more and better education will - in the long run - be a more successful policy than redistributing income or expanding public health care expenditures.

The question is thus: does education improve health? One theoretical explanation why more education might causally lead to better health is that education raises an individual's efficiency in health production, i.e. education raises the marginal productivity of inputs into health production (Grossman (1972)). Just like more educated people are more productive in labor market activities they are likely to be more productive in non-market activities, which include the production of their own health (or the health of their children). Better educated people would thus be healthier even if inputs were held fixed.

Another (complementary) theoretical explanation is that education changes the inputs into health production (through information) and thereby increases allocative efficiency. Most prominently, education might change health behaviors such as smoking or bad nutrition habits (Rosenzweig and Schultz, 1981; Cutler and Lleras-Muney, 2010). It has been shown that education is associated with substantially reduced smoking and obesity rates, two important causes of premature deaths (Mokdad et al. (2004)). For example, smoking rates in Germany are 48% among less educated men, as opposed to 20% among men with college degrees (Deutsches Krebsforschungszentrum (2004)). Education might improve health behaviors for several reasons. First, better educated people might be better informed about negative health consequences of smoking and overeating, either because they learned about these consequences in school, or because better educated people find it easier to obtain and evaluate such information (Kenkel, 1991; Nayga, 2000; de Walque, 2007). Second, education could also influence health behavior through higher income, different social environments, a different sense of control, or an impact of education on time preferences, e.g. because schooling focuses students' attention on the future (Fuchs (1982); Becker and Mulligan (1997)).

Despite many convincing arguments why there should be a causal effect of formal education on health, concerns about an interpretation of correlations as causal relationships remain. Most importantly, unobserved heterogeneity could drive both educational attainment at younger ages and health outcomes at older ages. For instance, Carneiro et al. (2007) find strong correlations between cognitive and non-cognitive skills measured at age 11 and measures of educational success such as having continued school beyond age 16 - reported at age 23 - or the highest educational qualification obtained - measured at age 43. Moreover, they find strong correlations between cognitive and non-cognitive skills measured at age 11 and the probability of smoking at age 16 (measured at age 16), and self-assessed health, depression and mental health problems in adulthood (all measured at age 42). This is evidence for the presence of - usually unobserved - variables driving the relationship between education and health.

The aim of the present paper is to estimate the causal effect of schooling on health exploiting variations in compulsory schooling laws in Western Germany. We follow the approach of Pischke and von Wachter (2005, 2008) who have used changes in years of compulsory schooling to estimate the causal effect of education on earnings. The authors found zero returns to schooling on wages. However, when estimating a reduced form regression of body mass index on the change in compulsory schooling laws they found a significant effect. Our objective is to take this line of research further and assess the non-monetary returns of education by considering other health outcomes and also important dimensions of health-related behavior such as smoking. There is now a number of papers exploiting reforms in mandatory schooling laws across different countries, for instance the United States (Lleras-Muney (2005)), the United Kingdom (Oreopoulos, 2006; Silles, 2009; Clark and Royer, 2010), or France (Albouy and Lequien (2009)). The results of these studies are mixed, showing negative effects of education on mortality (i.e. positive effects on health) in the US (Lleras-Muney (2005)) but not in France (Albouy and Lequien (2009)) or in the UK (Clark and Royer (2010)). In contrast, strong positive effects on self-rated health in the UK are found by Silles (2009).[1] Yet further studies for the UK, assessing the effect of compulsory schooling on morbidity using objective health measures such as blood pressure, BMI, or levels of inflammatory blood markers (Clark and Royer, 2010; Jürges et al., 2011) find no effects, while Powdthavee (2010) finds negative effects on hypertension.

Our study extends the existing literature by analyzing compulsory schooling laws in Germany to assess the causal effect of education in yet another

[1] Oreopoulos (2006) reports strong positive effect on self-rated health. However, a reanalysis of the data has basically reduced this effect to zero (see revised table 5 posted on the AER website).

economic environment where the economic returns to education at the margin of the educational reform were shown to be pretty low. In addition, our paper complements earlier work on the health effect of education in (Western) Germany, using the abolition of fees in academic track schools and academic track school constructions as different sources of exogenous variation in schooling (Jürges et al., 2011; Reinhold and Jürges, 2010).

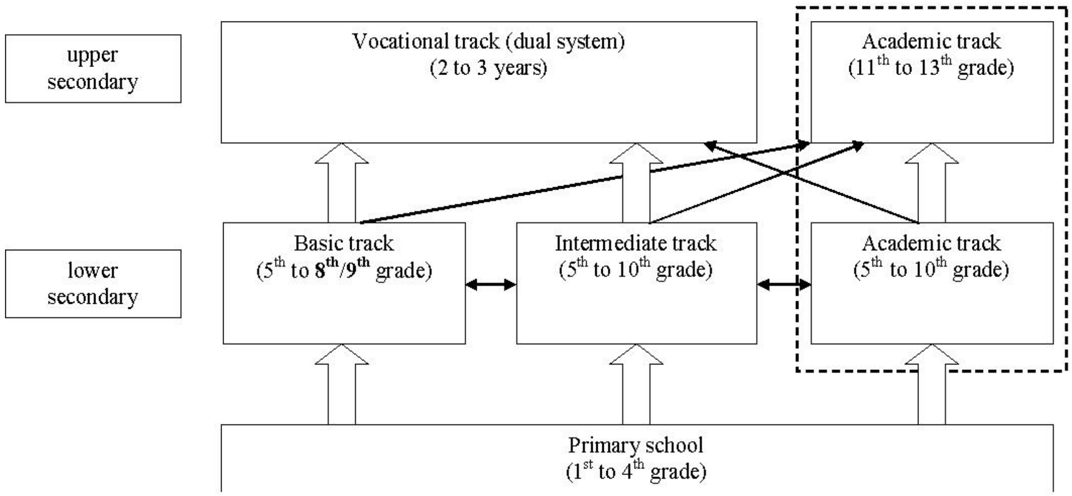
The paper is structured as follows. In the next section we will describe the German school system and the changes in compulsory schooling laws as our institutional background. Section 3 explains which data we use and shows descriptive statistics. Section 4 describes our identification strategy and its assumptions. Section 5 shows the results. Various robustness checks are reported in Section 6 and Section 7 gives a summary and discussion.

1.2 Institutional background

In Germany, the federal states are responsible for educational policy. Still, the educational systems in all German states are very similar and thus share many common features (see figure 1.1). Children generally enter primary school at the age of six. After (normally 4 years) in primary school, students attend one of three secondary school tracks which are in most cases taught at geographically separate schools. Hauptschule is the basic track leading to a basic school leaving certificate after grade 8 or 9. Realschule is a more demanding intermediate track which leads to a school leaving certificate after grade 10. Having finished school, both students from the basic track and the middle track usually start an apprenticeship or a school-based vocational training. Gymnasium is the academic track leading to a general university-entrance diploma (Abitur) after grade 12 or 13.[2] It is noteworthy that, after being allocated to one of the three secondary school tracks, students rarely move between tracks, so that the track choice at the age of 10 usually has a strong implication for the entire life course (Dustmann, 2004; Jürges and Schneider, 2007). The selection process itself depends on a mix of formal exams, grades in primary school, recommendations by the class teacher, and parental choice.

[2] In addition to the three common track types, some states also have comprehensive schools where all students are taught together (with internal tracking to some degree) or "middle schools", combining basic and intermediate track. Both types of school are relatively new and not relevant in our study.

Figure 1.1: Stylized description of the German school system at the time of the reforms



Since World War II, the German school system has undergone a number of reforms, among them the abolition of secondary school fees (Riphahn (2008)), or the large-scale expansion of academic track schools (Jürges et al. (2011)). The reform we try to exploit in the present paper is the prolonging of compulsory basic track schooling from eight to nine years. Since education policy is in the hands of state governments, this reform has taken place at different points in time. Table 1.1 (taken from Pischke and von Wachter (2005)) contains, by federal state, the first year when all students were supposed to stay in school for at least 9 years and the corresponding birth cohorts that were first affected by the reform. The first state to introduce a 9th grade in the basic track was Hamburg in 1949. The last state was Bavaria introducing a 9th grade in 1969. Interestingly, the data show a fairly consistent north-south pattern, with northern states being earlier reformers than southern states. Southern states tend to be politically more conservative and have a higher proportion of Catholics. Such political and cultural differences could give rise to differences in health and health behavior as well. In our analysis, we therefore control for level differences between states and also allow for state-specific trends. Note further that in contrast to education policy, health policy is the sole responsibility of the federal government. This is important because it lends credibility to our main identifying assumption of no correlation between state trends in education and health policies. In a similar vain we argue that due to nationwide public television health information (e.g. on the hazards of smoking or obesity) should be disseminated in all states at the same time and at a similar speed.

Table 1.1: Introduction of a 9th grade in the basic track of secondary school

	First year in which students were supposed to graduate no sooner than after nine years of compulsory schooling	First birth cohort affected by the change in compulsory schooling laws
Hamburg	1949	1934
Schleswig-Holstein	1956	1941
Bremen	1958	1943
Lower Saxony	1962	1947
Saarland	1964	1949
North Rhine-Westphalia	1967	1953
Hesse	1967	1953
Rhineland-Palatinate	1967	1953
Baden-Wuerttemberg	1967	1953
Bavaria	1969	1955

Source: Pischke and von Wachter, 2005

In some states, changes in compulsory schooling laws were undertaken simultaneously with other reforms. When compulsory schooling was raised to nine years, some German states also moved the start of the school year to early summer (cf Pischke (2007)). This implied the introduction of two short school years so that affected cohorts did in fact nominally complete 9 instead of 8 grades although they spent only one third of a complete school year longer in school compared to the previous cohorts unaffected by the reform. It is hard to assess how much more these cohorts learnt in comparison to previous cohorts. But potentially we underestimate the effect of the increase in compulsory schooling if we do not take this transient shortening of the school year into account, and our estimates would then be interpreted as lower bounds of the true effect. We address this problem thoroughly in our section on robustness checks.

1.3 Data and descriptive statistics

The data used in our study are taken from five years of the German Microcensus (1989, 1995, 1999, 2002, and 2003). The Microcensus is an annual, representative survey of one per cent of the households in Germany. Participation in the Microcensus is mandatory. However, the questions on health and health-related behavior are voluntary and are asked of a 45 percent random subsample only. The data contain, for each individual, sex and age, highest school degree and the current state of residence. In our analysis, we include cohorts born between 1930 and 1960 and currently living in the ten former West German states (excluding Berlin). We further restrict our analytical sample to individuals of German nationality and exclude individuals from the sample who have received their highest school degree in the former German Democratic Republic. Questions on health status (long-term illness and work-related disability), available only in the 2002 survey, were asked of working age individuals (16-65 years). We thus need to restrict our analysis to cohorts born between 1937 and 1960 when using the 2002 data on health status. We further exclude all individuals living in Hamburg from the 2002 survey since the change in compulsory schooling laws there first affected cohorts born in 1934 and thus there is no exogenous variation in schooling.

Direct information on years in school as such is not available in the Microcensus. But the data provide the highest secondary school degree attained by each respondent. We use this information, together with both the number of years usually taken to obtain a certain degree and the compulsory schooling laws in the respective state, to impute years of schooling in each state, cohort and secondary school type. A dummy variable indicating whether the compulsory schooling laws require 8 or 9 years of schooling will serve as an instrument for an individual's years of schooling. Individuals without any school degree are either assigned 8 or 9 years of schooling depending on the compulsory schooling laws in effect. Since geographical information is limited to state of residence at the time of the survey, we need to assume that a person has attended school in his state of residence. This raises concerns about the validity of the results of our analysis which we will address in a robustness check.

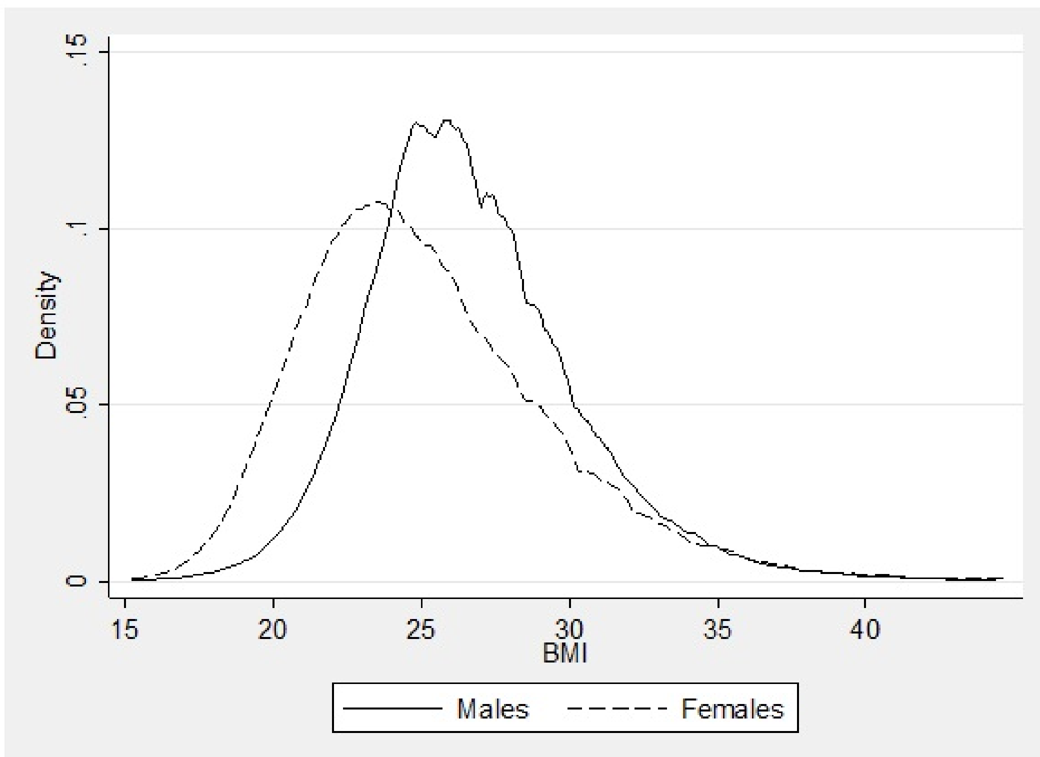
We measure health and health-related behavior of the respondents by a number of variables. Current health is measured by an indicator variable whether someone suffers from long-term illness (we do not have information on specific diagnoses) and whether this illness limits the type or amount of work they can do (work disability). As mentioned above, this information is only available in 2002 and only for working age individuals. Another measure of current health is the body mass index (BMI), derived from self-reported information on weight and height in the 1999 and 2003 surveys. We used BMI to construct two dummy variables indicating overweight (BMI greater than 25) and obesity (BMI greater than 30). Note that our definition of overweight includes obesity. Also note that BMI, overweight or obesity can also be interpreted as indicators of future health because they are significant risk-factors for

Table 1.2: Sample means (standard deviation in parentheses, number of observations below)

Variable	Sex	**Microcensus 2002** Cohorts born between 1937 and 1960	**Microcensus 1999, 2003** Cohorts born between 1930 and 1960	**Microcensus 1989, 1995, 1999, 2003** Cohorts born between 1930 and 1960
Age	Men	52.48	54.76	50.13
		(7.10)	(9.10)	(10.27)
		24,099	49,843	122,178
	Women	52.63	55.19	50.57
		(7.11)	(9.17)	(10.31)
		24,624	50,560	124,674
Years in school	Men	9.63	9.50	9.42
		(1.82)	(1.84)	(1.80)
		23,128	48,640	119,461
	Women	9.39	9.20	9.17
		(1.62)	(1.59)	(1.55)
		23,479	49,225	121,541
Long-term illness	Men	0.20	-	-
		(0.40)		
		24,099		
	Women	0.15	-	-
		(0.36)		
		24,624		
Work disability	Men	0.17	-	-
		(0.37)		
		23,791		
	Women	0.13	-	-
		(0.33)		
		23,369		
BMI	Men	-	26.70	-
			(3.66)	
			49,843	
	Women	-	25.25	-
			(4.36)	
			50,560	
Overweight	Men	-	0.66	-
			(0.47)	
			49,843	
	Women	-	0.46	-
			(0.50)	
			50,560	
Obesity	Men	-	0.16	-
			(0.36)	
			49,843	
	Women	-	0.13	-
			(0.34)	
			50,560	
Ever smoked	Men	-	-	0.63
				(0.48)
				122,178
	Women	-	-	0.39
				(0.49)
				124,674
Currently smoking	Men	-	-	0.36
				(0.48)
				122,178
	Women	-	-	0.24
				(0.43)
				124,674
Quitted smoking	Men	-	-	0.44
				(0.50)
				77,566
	Women	-	-	0.38
				(0.49)
				48,501

future health problems, e.g. diabetes or cardiovascular disease. We measure health behavior by information on smoking. In 1989, 1995, 1999, and 2003, individuals were asked whether they were smoking currently and, if not, whether they had ever smoked. We use this information to identify quitters. Information on number of cigarettes smoked is only reported in rather broad bands. Analyses based on this measure delivers qualitatively similar results compared to our results using just binary information on smoking status. We generally pool all survey years which contain the same dependent variables, because each survey year represents an independent random sample.

Figure 1.2: Kernel density estimate of BMI



Descriptive statistics are presented in table 1.2. There are roughly 25,000 individuals per sex in the 2002 "current health" sample. These individuals were born between 1937 and 1960 and were on average 52 to 53 years old in 2002. Mean years in school were 9.6 years for men and 9.4 years for women. 20% of men and 15% of women were suffering from a long term illness. 17% of men and 13% of women suffered from a long-term illness that limited the amount or type of work they could do, i.e. they were work disabled. In the "BMI sample", interviewed in 1999 and 2003, there are about 50,000 observations per sex. Average BMI was 26.7 for men and 25.3 for women. 66% of men were overweight or obese compared to only 44% of women. Obesity rates were 16% for men and 13% for women. These sex differences can also be seen when looking at the kernel density estimates (figure 1.2) showing that the distribution is shifted to the right for men. For both sexes, one can see that a lot of the probability mass is concentrated in the BMI region around 25. For men, it even peaks at around those values.

For the analysis of smoking behavior, we have the largest samples with more than 120,000 observations for each sex. Overall, 63% of men and 39% of women reported having ever smoked. Current smoking rates are 36% for men and 24% for women. This implies that 44% of men (38% of women) in our sample who ever smoked have stopped smoking.

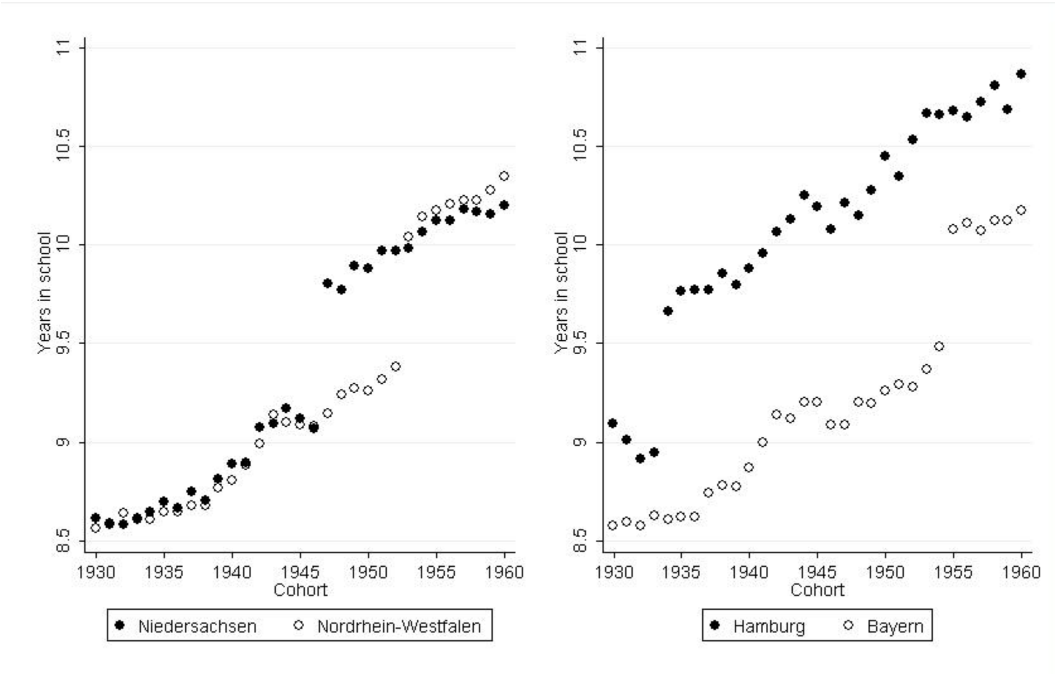
Table 1.3: Distribution of leaving certificates in percent, by birth cohorts

		1930	1940	1950	1960
Basic Track	Men	76.8	69.1	58.8	47.7
	Women	82.1	73.0	63.0	40.7
Intermediate	Men	12.3	14.6	17.2	22.7
	Women	12.9	19.8	22.4	35.5
Academic	Men	10.9	16.3	23.9	29.6
	Women	4.9	7.2	14.6	23.8

Table 1.3 shows the proportion of individuals in the birth cohorts 1930, 1940, 1950, and 1960 who graduated from each of the three secondary school tracks. These data clearly show the educational expansion in West Germany after the Second World War. More than three quarters of the population born in 1930 received at most basic track schooling. This proportion has decreased to less than 50% among those born in 1960. Moreover, the decline was much larger for women than for men. Both intermediate and academic track education has increased correspondingly. As a consequence, there is substantial variation across birth cohorts and states with respect to mean years of schooling (see table 1.3 for cohorts and figure 1.3 for selected states).

Figure 1.3 shows upward trends in average years of levels for four selected Western German states (graphs for other states look very similar), from values below 9 years for the 1930 birth cohorts to values above 10 years for the 1960 birth cohorts. In addition to those time trends for the whole period, one can also see jump of about 0.6 years in mean years of schooling at the time the change in mandatory schooling laws took effect. There are some more noteworthy points in figure 1.3. For instance, Lower-Saxony and North Rhine-Westphalia were

Figure 1.3: Mean number of years of schooling by birth cohort, selected federal states



very much on the same "expansion path" until Lower-Saxony lengthened the basic track, and North Rhine-Westphalia caught up after prolonging mandatory schooling 6 years later. Further, the pictures illustrate the need to control for state-specific trends, as for example Schleswig-Holstein has a flatter increase in average educational attainment than Baden-Wuerttemberg (and other states). In our basic specification, we will thus control for state-specific linear trends in addition to (common) time fixed effects to control for these different paths of the educational expansion across states.

1.4 Empirical strategy

We estimate both OLS and IV models. The former are estimated for descriptive purposes while the latter are given a causal interpretation. All equations are estimated for men and women separately and with robust standard errors which are clustered by state and year of birth. Our OLS model linking the health outcome to an individual's years of schooling is specified as:

$$H_i = \beta_0 + \beta_1 S_i + \beta_2 age_i^2 + \mu_{state} + \delta_{state} \times cohort_i + \nu_{cohort} + \eta_{survey} + \epsilon_i \quad (1.1)$$

where H_i is a variable indicating some health outcome for individual i. S_i denotes individual i's years of schooling. μ_{state}, ν_{cohort}, and η_{survey} are fixed effects accounting for heterogeneity between federal states, between birth cohorts and between survey years. In addition, we add linear cohort-state interaction effects ($\delta_{state} \times cohort_i$) controlling for state-specific deviations from the common (non-parametrically estimated) nationwide trend captured by the cohort dummies (thus if k is the number of states, we estimate k-1 state fixed effects and k-1 state specific cohort effects). Note that we cannot include a linear age effect because it is perfectly collinear with cohort and survey year combined. Yet we are able to include quadratic age term in order to address potentially non-linear age effect. When using the 2002 wave only, the survey-specific fixed effects and the quadratic in age are excluded from the right-hand side of equation (1).

The method of instrumental variables - using changes in compulsory schooling laws discussed above as instruments - allows dealing with the potential endogeneity of education in equation (1) and thus allows estimating a causal effect of education on health and health-related behavior. Our identification rests on the assumption that - conditional on covariates - compulsory schooling laws or changes therein are uncorrelated with health outcomes and health behaviors, except through their effect via education. Technically, we assume that state education policies are independent of any unobserved determinants of health in the second stage (health) equation.

One particular concern is that state-specific trends in health or health behavior may be correlated with both the change in compulsory schooling laws and unobserved characteristics of the individuals that are themselves health determinants. If this was the case, the association between the instrument and the health outcome would not only be due to the association between the instrument and years of schooling. For example, there may be macroeconomic factors such as higher per capita growth which positively affect both the income of (parental) households and a state's resources which are available for reforms of the schooling system (such as financing one more school year for more than half of each cohort). If schools are financed by income taxes, there might also be a negative contemporaneous correlation between net parental household income and education expenses. In any case, as long as parental household income affects health (Case et al., 2002; Reinhold and Jürges, 2010), state-specific trends in economic development may compromise the validity condition. For this reason, we estimate the IV-model controlling for state-specific linear trends. These trends should also capture effects of other measures of public policy which might be implemented together with the change in compulsory schooling laws. In additional robustness checks, we also investigate the robustness of our results to the inclusion of state-specific quadratic and cubic trends.

We estimate the following first stage equation:

$$S_i = \gamma_0 + \gamma_1 Z_i + \gamma_2 age_i^2 + \rho_{state} + \kappa_{state} \times cohort_i + \tau_{cohort} + \lambda_{survey} + \omega_i \quad (1.2)$$

where S_i denotes again years of schooling. Z_i is an instrumental variable indicating whether compulsory schooling laws require individual i to stay at least 8 or 9 years in school. When constructing the schooling variable S_i, we assigned 8 or 9 years of schooling to the basic track according to the applicable compulsory schooling laws. Thus, the association between the instrumental variable and the schooling variable (γ_1) arises by construction. ρ_{state}, τ_{cohort} and λ_{survey} are fixed effects accounting for heterogeneity between federal states, birth cohorts and survey years and κ_{state} represents a state specific linear cohort trend. Inserting years of schooling predicted from equation (2), $\hat{S}_i$, into equation (1) yields the second stage equation:

$$H_i = \beta_0 + \beta_1 \hat{S}_i + \beta_2 age_i^2 + \mu_{state} + \delta_{state} \times cohort_i + \nu_{cohort} + \eta_{survey} + \epsilon_i \quad (1.3)$$

When using the 2002 wave only, the survey-specific fixed effects and the quadratic in age are excluded from the right-hand side of equations (2) and (3). There are two concerns about the interpretation of our parameter of interest β_1 in equation (3). First, it is possible that the change in factual school attendance has not occurred sharply when a 9th grade of compulsory schooling was introduced. Pischke and von Wachter (2005) show that the introduction of reform was somewhat spread out and did not occur sharply. As a consequence, years of schooling may be understated for some students in basic track shortly before the introduction of 9th grade and overstated for some students in the basic track shortly after the introduction of the 9th grade. The "true" first stage estimate accounting for this type of one-sided measurement error may thus be lower than our estimate of γ_1. Our IV-estimates would then suffer from attenuation bias and our results have to be interpreted as lower bounds. Back-of-the-envelope calculations based on numbers provided in Pischke and von Wachter (2005, figure 3) suggest that the true effects are not more than 15% larger than the estimates presented in our paper.[3]

Second, to allow for heterogeneous effects of education on health, we need to assume monotonicity, i.e. we need to rule out that after the lengthening of mandatory schooling some students decide to go to school shorter than they would without the reform (i.e. we assume that there are no "defiers"). In our case it seems plausible to assume that individuals with already high educational attainment (in intermediate or academic track) do not get less schooling because the mandatory years in basic track are raised by one year. This assumption would be violated if individuals chose a less demanding track or dropped out of school in reaction to the raise in years of compulsory schooling. We assess in a further robustness check whether this scenario is plausible. With heterogeneous causal effects and when the monotonicity assumption holds, we identify the effect of education only on compliers, i.e. individuals whose number of years in school is affected positively by the reform (Angrist and Imbens, 1995), which also includes individuals who decide to stay longer than the mandatory 9 years. Moreover, since the number of years in school is a multi-valued treatment, the IV estimates are equal to a weighted average of all complier-specific causal effects.

1.5 Results

1.5.1 OLS results

Long-term illness and work disability. OLS regression results are shown in table 1.4. The OLS coefficients indicate a significant association between current health and education in both sexes. It is, however, stronger for men than for women. One more year in school is linked with a reduction in the likelihood of suffering from a long-term illness of 2.9 percentage points for men but only 1.2 percentage points for women. The association between years of schooling and long term illness is of substantive relevance considering that the overall prevalence rate of long term illness is 20% for men and 15% for women. The association between years in school and the probability of being work disabled is very pronounced and significant for both sexes, and it has about the same

[3] On the other hand, reduced form (or "intention-to-treat") estimates, which do not suffer from any measurement problem in actual years of schooling but which measure the reform effect on compliers and always-takers combined, are about one third smaller than IV-estimates (see table 1.9 in the appendix).

magnitude as the association of schooling and long-term illness.

Table 1.4: OLS results

Dependent variable		Sex	**Coefficient on years in school**	**Standard Error**	**Number of observations**
Current health	Long-term illness	Men	-0.029***	0.001	23,128
		Women	-0.012***	0.002	23,479
	Work disability	Men	-0.029***	0.001	22,858
		Women	-0.013***	0.001	23,266
Weight	BMI	Men	-0.271***	0.010	48,640
		Women	-0.465***	0.013	49,225
	Overweight	Men	-0.031***	0.001	48,640
		Women	-0.047***	0.002	49,225
	Obesity	Men	-0.018***	0.001	48,640
		Women	-0.020***	0.001	49,225
Smoking	Ever smoked	Men	-0.026***	0.001	119,461
		Women	-0.005**	0.002	121,541
	Currently smoking	Men	-0.032***	0.001	121,318
		Women	-0.021***	0.002	124,314
	Quitted smoking	Men	0.027***	0.001	76,037
		Women	0.041***	0.002	47,551

Note: Standard errors clustered on cohort*state level. * p<10 percent, ** p<5 percent, *** p<1 percent. All regressions include fixed effects for year of birth, state of residence, and an interaction of state of residence and a linear cohort trend. If several survey years have been pooled, fixed effects for survey year and the quadratic in age are included. Regressions with weight outcomes include height as additional control variable.

BMI, overweight and obesity. The OLS estimates also indicate a significant association between years in school and body weight. Here, the association is much stronger for women than for men. One more year in school is linked with a decrease in BMI of 0.27 kg/m2 for men and 0.47 kg/m2 for women. In terms of overweight and obesity this means that each additional year in school is linked with a 3.1 (1.8) percentage point lower probability of being overweight (obese) for men, and a 4.7 (2.0) percentage point lower probability of being overweight (obese) for women. Again, these are fairly strong associations.

Smoking. Smoking behavior is also strongly related to schooling, but more so for men than for women. The OLS estimates show that one more year of schooling is linked with a 2.6 percentage point lower proportion of men who ever smoked but only a 0.5 percentage point lower proportion of women who ever smoked. With respect to current smoking, the estimates suggest that one more year of schooling is associated with a decrease in the likelihood of being a smoker of 3.2 percentage points for men and of 2.1 percentage points for women. Further, the OLS estimates suggest a stronger association between years of schooling and the probability of quitting smoking for women than for men. One more year of schooling is associated with an increase in the probability of having quitted smoking of 4.1 percentage points among women and of 2.7 percentage points among men.

1.5.2 Instrumental variables results

Table 1.5 shows the relevant first and second stage coefficients from our IV estimation. As indicated by the first stage results, mandatory schooling laws have high explanatory power with respect to years of schooling. The introduction of a compulsory 9th grade leads to an average increase of about 0.6 years in school. We use the Kleibergen-Paap weak identification test to assess the strength of our instruments. All test statistics are above 157 giving no rise to concerns about weak instrument problems.[4] It is also interesting to investigate for whom we identify the effect of education on health. Following Daron Acemoglu (2000), we estimate the average causal response weights, i.e. the weight different complier groups contribute to our estimate. Not surprisingly, we find that these weights are concentrated on individuals having nine years of schooling (96.5%) and much less weight on individuals with 10 years of schooling (2.3%) or even higher levels of schooling (1.2%). Thus, we largely identify an effect for individuals in the basic track getting nine instead eight years of schooling.

Long-term illness and work disability. IV estimates of the effect of schooling on health outcomes differ substantially between men and women. For men, we find a large and significant causal effect of years of schooling on health. Our results indicate that one more year of schooling reduces the likelihood of suffering from a long term illness by 4.1 percentage points for men (5% significance level). Furthermore, the estimates for men suggest that one more year of schooling leads to a reduction in the likelihood of work disability of 3.2 percentage points, which is close to the OLS result. In contrast to men, women appear not to gain from more schooling in terms of current health. The IV coefficients for women now have a positive sign but have become insignificant.

One possible explanation for the difference between men and women in the effect of education on work disability is that men with low levels of education tend to work more often in sectors where hard physical labor is performed, in particular in those older cohorts where labor force participation of women was

[4]It has to be taken into account that the German Microcensus does not contain a direct measure of years of schooling. Since we assigned 8 or 9 years of schooling to the basic track according to the applicable compulsory schooling laws, the strong association between the instrumental variable and the schooling variable arises largely by construction.

Table 1.5: Results of 2SLS regressions

Dependent variable		Sex	**First Stage coefficient on instrument**	**Second stage coefficient on years in school**	**Number of observations**
Current health	Long-term illness	Men	0.656*** (0.054) 160.3	-0.041** (0.017)	23,128
		Women	0.579*** (0.047) 161.8	0.010 (0.017)	23,479
	Work disability	Men	0.655*** (0.055) 157.7	-0.032** (0.015)	22,858
		Women	0.580*** (0.047) 160.5	0.021 (0.016)	23,266
Weight	BMI	Men	0.595*** (0.040) 238.0	-0.301** (0.121)	48,640
		Women	0.663*** (0.033) 380.2	-0.194 (0.133)	49,225
	Overweight	Men	0.595*** (0.040) 238.0	-0.030** (0.015)	48,640
		Women	0.663*** (0.033) 380.2	-0.031** (0.015)	49,225
	Obesity	Men	0.595*** (0.040) 238.0	-0.030** (0.014)	48,640
		Women	0.663*** (0.033) 380.2	-0.004 (0.010)	49,225
Smoking	Ever smoked	Men	0.615*** (0.025) 620.8	-0.011 (0.011)	119,461
		Women	0.655*** (0.021) 952.3	0.010 (0.010)	121,541
	Currently smoking	Men	0.616*** (0.025) 631.5	-0.005 (0.010)	121,318
		Women	0.654*** (0.021) 975.1	-0.001 (0.009)	124,314
	Quitted smoking	Men	0.668*** (0.030) 536.2	-0.005 (0.011)	76,037
		Women	0.690*** (0.031) 507.8	0.015 (0.012)	47,551

Note: Clustered standard errors (on cohort*state level) in parentheses. In addition, we report Kleibergen-Paap rK Wald F statistics in italics. All regressions include fixed effects for year of birth, state of residence, and an interaction of state of residence and a linear cohort trend. If several survey years have been pooled, fixed effects for survey year and the quadratic in age are included. Regressions with weight outcomes include height as additional control variable. * p<10 percent, ** p<5 percent, *** p<1 percent.

much lower. For this reason, an additional year of schooling could be of particular relevance for the working conditions of men if it pushes men from blue to white collar jobs. Education is associated with non-manual labor occupations where dangers to physical health through exposure to injury, dangerous chemicals or adverse ergonomic conditions are reduced (World Health Organization (2003)). Furthermore, it could be the case that an individual with given health problems, as for instance back problems, would be able to work in a white-collar job but not in a blue-collar job anymore. If this was the case, we would also expect more cases of work disability for blue-collar workers. To investigate this hypothesis further, we included a set of occupational dummy variables in our regressions for long-term illness or work disability.[5] We found that the size of the coefficient on years of schooling was reduced by roughly 25% for men (detailed results not shown). Part of the education effect on health thus seems to work through occupational choice. We also found that the effect of education on working in a white-collar job was smaller for women than for men. Therefore, it seems that gender differences in the effect of education on long term illness and work disability can partly be explained through the choice of occupational sector.

BMI, overweight, and obesity. For men's BMI, we find a significant effect of about the same size as the OLS results. One more year in school decreases male BMI by 0.3 kg/m2. Among women, the effect approximately halves in size and becomes insignificant. Each year in school significantly reduces the probability of being overweight by 3.0 percentage points for men and 3.1 percentage points for women. Our effect sizes are consistent with the estimates for BMI in Pischke and von Wachter (2008), table 3) who find a reduced form coefficient on mandatory schooling laws of -0.16 kg/m2 in the pooled sample of men and women. The relatively small effect on BMI and the strong effect on overweight can be explained by the fact that a lot of the density mass is concentrated right around the critical threshold of a BMI of 25 (see figure 1.3). The effect size is again in the range of the OLS coefficients for men but somewhat smaller for women. With respect to obesity, we obtain significant IV coefficients only for men but not for women.

Smoking behavior. In contrast to our health outcome measures and in contrast to the corresponding OLS results, our IV estimates do not suggest any significant effect of education on smoking behavior. This holds for having ever smoked, being a current smoker, and for having quitted smoking and for both sexes. With one exception, the IV coefficients have become much smaller

[5]Note that we did not include occupational dummies in our main regressions because occupation is itself determined by education.

than the OLS estimates and some coefficients even change signs. We checked whether the effect of education on smoking behavior is so different from the other health variables studied here because the data come from a longer and on average earlier period by restricting the sample to the survey years 1999 and 2003. IV estimates remained essentially zero.

1.6 Robustness checks

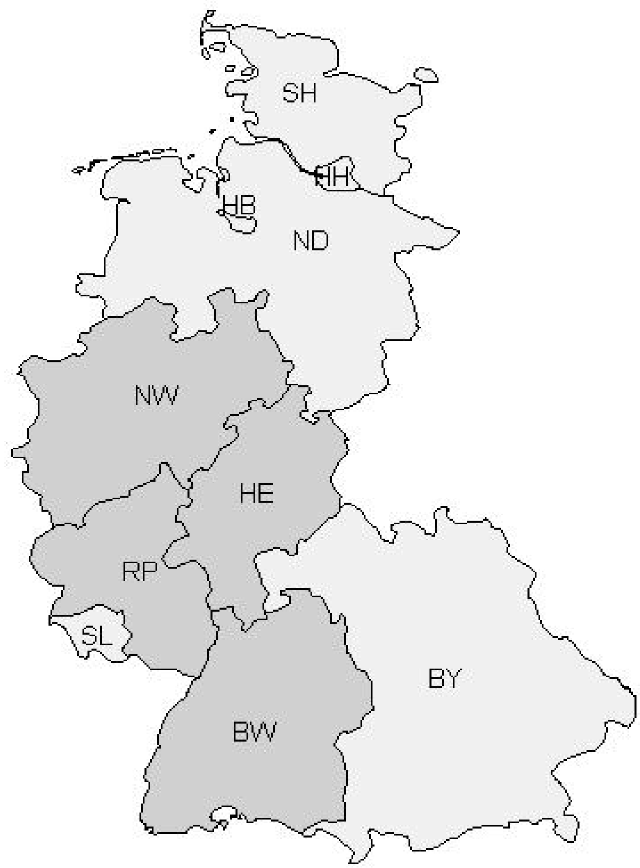
We now address several concerns regarding our empirical strategy as mentioned before. First, we only have information on the current state of residence, but not on the state in which the individual has finished schooling. We thus restrict our sample to four contiguous states that have changed mandatory schooling laws at the same time. Second, we relax our assumption that the secular trends in health in Western Germany can be captured by (a common) cohort fixed effects plus a linear state-specific trends by estimating models adding also quadratic and cubic state-specific trends. Third, we investigate whether the mandatory schooling reforms had additional effects on track choice and dropping out behavior - as explained below this will give us some idea about the plausibility of the monotonicity assumption. And fourth, concurrent with raising compulsory schooling to nine years, the start of the school year was shifted from Easter to the end of summer in all German states. This was done by introducing short school years reducing the time spent in school by up to two thirds of a school year in some states (cf. Pischke (2007)). We account for this problem by recoding our endogenous variable years of schooling accordingly.

1.6.1 Restricting the sample to four contiguous states

A major concern about the validity of our results stems from the fact that geographical information is limited to state of residence at the time of the survey. Thus, we have to assume that a person has attended school in his current state of residence when constructing the instrumental variable as well as the schooling variable. This introduces a measurement error in both the instrument and the endogenous regressor, schooling. This is one source of attenuation bias discussed in Pischke and von Wachter (2005, 2008). We address this concern here by restricting our sample to four contiguous states that have reformed mandatory schooling simultaneously (see figure 1.4).

Baden-Wuerttemberg, Rhineland-Palatinate, North Rhine-Westphalia, and Hesse introduced the 9th grade of compulsory schooling in 1967. In 1970, these

Figure 1.4: Map of West Germany showing the four contiguous states that lengthened compulsory schooling in the same year (NW=Northrhine-Westphalia, HE=Hesse, RP=Rhineland-Palatinate, BW=Baden-Wuerttemberg)



four states represented 57% of the West German population. Since they do not border on a city state and contain large cities themselves, migration between the four states and other German states should be a minor issue. Furthermore, migration occurring within those four states does not matter since the 9th grade has been introduced in the same year. In a first step, we analyzed the same birth cohorts as previously (1930-1960), simply restricting our sample to the four contiguous states. When restricting the sample we obtain qualitatively very similar results compared to the whole sample (detailed results not shown), indicating that attenuation bias due to migration between the states is not a major driver of our results.

The effect of the reform can in some sense be more satisfactorily estimated if we restrict our analysis to cohorts who are just affected and just not affected by the introduction of the 9th year of compulsory schooling in the basic track. Including birth cohorts born between 1930 and 1960 was useful in the original analysis (including all federal states) because school reforms were dispersed over a larger time period. The downside is that one might be concerned about the ability to adequately capture the time trends even in the fairly flexible specification that we use. In the four-state sample, we can further restrict the sample to cohorts to the last cohort for whom the change in compulsory schooling laws was not yet applicable and the first cohort who was supposed to attend 9 years of schooling in the basic track. One problem with this approach, however, is that for those four states the raise in compulsory schooling laws takes place at the same time as the move of the start of the school year to the summer. This rescheduling was done by introducing short school years. While completing nine grades instead of eight the cohorts immediately affected by this rescheduling spent only around 8 and a third years in school which is not much more than the previous cohorts. We address this question separately below. In any case, we can interpret our coefficient estimates as a lower bound for the effect of an additional school year.

Our identification rests on the assumption that cohorts just before and just after the reform took effect are very similar and differ only with respect to the treatment. In this case, we do not rely on estimating unobserved trends on the state level. In other words, we assume that the potential no-treatment outcomes of the two cohorts are identical. Technically, this means that we set cohort fixed effects to zero. We thus modify our IV estimation approach as follows: The first stage equation becomes:

$$S_i = \gamma_0 + \gamma_1 Z_i + \gamma_2 age_i + \gamma_3 age_i^2 + \rho_{state} + \lambda_{survey} + \omega_i \qquad (1.4)$$

The cohort-specific fixed-effects have been excluded from the right-hand side of equation (2). On the other hand, a linear age effect has now been included since it is not captured anymore by cohort-specific and survey-specific fixed effects.

The second stage equation becomes:

$$H_i = \beta_0 + \beta_1 \hat{S}_i + \beta_2 age_i + \beta_3 age_i^2 + \mu_{state} + \eta_{survey} + \epsilon_i \qquad (1.5)$$

When we estimate the effect on health outcomes that are only present in a single survey year of the German Microcensus, the survey-specific fixed effects, the linear age effect and the quadratic in age are excluded from the right-hand side of equations (4) and (5). The results are shown in table 1.6. Because of the low number of clusters, we do not use clustered standard errors in the restricted sample but rely on conventional heteroskedasticity-robust standard errors.

Table 1.6: Robustness check (sample restricted to Baden-Wuerttemberg, Rhinland-Palatinate, Hesse and Northrhine-Westphalia)

Dependent variable		Sex	**OLS**	**2SLS First stage**	**2SLS Second stage**	**N obs.**
Current health	Long-term illness	Men	-0.028***	0.589***	-0.081**	1,097
			(0.006)	(0.111)	(0.037)	
		Women	-0.011*	0.716***	-0.013	1,199
			(0.006)	(0.093)	(0.028)	
	Work disability	Men	-0.027***	0.592***	-0.074**	1,095
			(0.005)	(0.111)	(0.034)	
		Women	-0.010*	0.721***	-0.014	1,194
			(0.005)	(0.093)	(0.025)	
Weight	BMI	Men	-0.118**	0.512***	-0.721	1,874
			(0.050)	(0.125)	(0.526)	
		Women	-0.412***	0.692***	-0.751*	1,772
			(0.059)	(0.114)	(0.438)	
	Overweight	Men	-0.010*	0.512***	-0.061	1,874
			(0.006)	(0.125)	(0.067)	
		Women	-0.041***	0.692***	-0.064	1,772
			(0.007)	(0.114)	(0.048)	
	Obesity	Men	-0.013***	0.512***	-0.020	1,874
			(0.004)	(0.125)	(0.049)	
		Women	-0.015***	0.692***	-0.072**	1,772
			(0.004)	(0.114)	(0.032)	
Smoking	Ever smoked	Men	-0.033***	0.507***	0.021	4,525
			(0.004)	(0.080)	(0.041)	
		Women	-0.018***	0.647***	-0.064*	4,447
			(0.005)	(0.069)	(0.034)	
	Currently smoking	Men	-0.047***	0.512***	0.023	4,585
			(0.004)	(0.079)	(0.042)	
		Women	-0.036***	0.650***	-0.037	4,513
			(0.004)	(0.069)	(0.031)	
	Quitted smoking	Men	0.041***	0.640***	-0.018	3,029
			(0.005)	(0.092)	(0.040)	
		Women	0.047***	0.677***	-0.009	2,346
			(0.006)	(0.093)	(0.042)	

Note: Robust standard errors in parentheses. All regressions include fixed effects for state of residence. If several survey years have been pooled, fixed effects for survey year and the quadratic in age are included. Regressions with weight outcomes include height as additional control variable. * $p<10$ percent, ** $p<5$ percent, *** $p<1$ percent.

The IV estimates for the restricted sample largely support our earlier findings - even though there are some differences. Naturally, the estimates are less precisely estimated when using the restricted sample. The estimated effect of education on long-term illness and work disability among men has increased substantially compared to the estimates for the unrestricted sample. For women, the point estimates have turned negative but they remain statistically insignificant. We also find substantially larger effects of years in school on BMI. Compared to the point estimates in the full sample the effects have more than doubled in size for men. For women they have more than tripled. Statistical significance is weak, however, which may be a result of the small size of the restricted sample. Similarly, despite substantially larger point estimates, we find a statistically significant effect of years in school on the probability of being overweight or obese only for women - which is in contrast to our results for the unrestricted sample.

Finally, considering smoking behavior, most effects remain insignificant although there are some changes with respect to the effect sizes. However, we find a significant effect of years of schooling on the probability of ever having smoked for women (at the 10% significance level) suggesting that among women, one more year of schooling causes a decrease in the probability of ever having smoked of 6.4 percentage points. Compared to the prevalence rate of 52.4% in the restricted sample, the estimated effect is relatively large.

1.6.2 Allowing for more flexible state-specific trends

We now analyze whether controlling for state-specific linear cohort trends in addition to (common) cohort fixed effects is sufficient to control for unobserved trends at the state level which might be correlated with health and with the decision to institute mandatory schooling laws. one becomes more comfortable with the results. We present additional results including state specific quadratic or cubic trends in table 1.7.

These results indicate that for men our results are very robust to the inclusion of more flexible state-specific trends. For long-term illness and work disability the coefficients maintain their significance levels and even slightly increase in size indicating a strong protective effect of education for men. For women, however, we still find basically a zero effect for long-term illness and work disability. These results are mirrored for BMI, and the probabilities of being overweight or obese. For men, the coefficients on education do not change much as one models the state-specific trends more flexible. However, for women, we now find a zero effect on weight problems. For smoking, we still do not find any significant results. These robustness checks suggest that there is only weak evidence for a causal effect of education for women's health in our sample when using our instrument (weak meaning it relies on more restrictive assumptions). However, our findings for men are strengthened.

Table 1.7: Robustness check (sample restricted to Baden-Wuerttemberg, Rhinland-Palatinate, Hesse and Northrhine-Westphalia)

Dependent variable		Sex	**+linear state trends**	**+quadratic trends**	**+cubic trends**	**N obs.**
Current health	Long-term illness	Men	-0.041**	-0.039**	-0.045**	23,128
			(0.017)	(0.019)	(0.018)	
		Women	0.010	0.013	0.024	23,479
			(0.017)	(0.024)	(0.027)	
	Work disability	Men	-0.032**	-0.037**	-0.045***	22,858
			(0.015)	(0.017)	(0.017)	
		Women	0.021	0.036	0.040	23,266
			(0.016)	(0.022)	(0.025)	
Weight	BMI	Men	-0.301**	-0.306**	-0.356**	48,640
			(0.121)	(0.136)	(0.146)	
		Women	-0.194	-0.050	0.149	49,225
			(0.133)	(0.150)	(0.168)	
	Overweight	Men	-0.030**	-0.034**	-0.039**	48,640
			(0.015)	(0.015)	(0.016)	
		Women	-0.031**	-0.016	-0.005	49,225
			(0.015)	(0.016)	(0.020)	
	Obesity	Men	-0.030**	-0.028*	-0.021	48,640
			(0.014)	(0.015)	(0.016)	
		Women	-0.004	0.007	0.023*	49,225
			(0.010)	(0.012)	(0.013)	
Smoking	Ever smoked	Men	-0.011	-0.012	-0.004	119,461
			(0.011)	(0.013)	(0.014)	
		Women	0.010	0.014	0.015	121,541
			(0.010)	(0.011)	(0.011)	
	Currently smoking	Men	-0.005	0.005	-0.001	121,318
			(0.010)	(0.012)	(0.014)	
		Women	-0.000	0.001	0.004	124,314
			(0.009)	(0.010)	(0.012)	
	Quitted smoking	Men	-0.005	-0.019	-0.006	76,037
			(0.011)	(0.013)	(0.014)	
		Women	0.015	0.015	0.011	47,551
			(0.012)	(0.015)	(0.017)	

Note: Robust standard errors in parentheses. All regressions include fixed effects for state of residence. If several survey years have been pooled, fixed effects for survey year and the quadratic in age are included. Regressions with weight outcomes include height as additional control variable. * p<10 percent, ** p<5 percent, *** p<1 percent.

1.6.3 Track choice and drop out

Next, we investigate whether the mandatory schooling reforms had additional effects on track choice and dropping out behavior. The sign of such effects is ex ante unclear. As shown previously, our instruments mainly identify the effect of schooling for individuals who acquire now 9 instead of 8 years of schooling. One crucial assumption in models with treatment effect heterogeneity is monotonicity, i.e. that all individuals react to the instruments in the same way (Angrist and Imbens (1995)). In our context monotonicity implies that individuals never prefer less education, e.g. they (weakly) prefer 10 years of schooling (feasible only in intermediate track before and after the reform) to 9 years (in basic track, after the reform) to 8 years (in basic track, before the reform). Extending the basic track by one year would thus push students in basic track from attending 8 to attending 9 years, while the choices of intermediate track students would remain unaffected (intermediate track students just do not care whether the basic track takes 8 or 9 years).

Can we imagine a situation in which the monotonicity assumption would be violated, i.e. where students get less schooling after the reform? "Defiers" in the sense that lengthening the basic track pushes students from attending 9 to attending 8 years in basic track cannot exist if the reform is enforced by law. However, for some the schooling choice before the reform was between 8 years in basic track and 10 years in intermediate track, and the choice after the reform was between 9 years in basic track and 10 years in intermediate track. Thus if someone's optimal number of years of education was about 9 years, prolonging basic track by one year would open up the opportunity to optimize education decisions by reducing schooling from 10 years in intermediate track to 9 years in basic track, and therefore also changing the track and the ensuing school leaving certificate. Yet other students may not complete school in reaction to the reform, i.e. although they stay one year longer in basic track, they leave without a certificate. Hence it would be reassuring if we found no effect of compulsory schooling laws on track choice, or that students attend a more demanding track in reaction to compulsory schooling laws, and no effect on "drop-out" rates.[6]

Another reason to look at track choice is to assess the potential role of peer effects. For instance, Jürges et al. (2011) look at the construction of new academic track schools increasing years of schooling mainly by increasing the proportion of students attending academic track schools. These shifts, however, also change the composition of students within the different school tracks. If we found only small effects of changes in compulsory schooling in basic track on track choice then the reforms would also likely not affect the composition of peers in the different school tracks.

To investigate track choice, we created a new dummy for students who graduated from the middle or academic track, and regressed this dummy on the same explanatory variables as in the previous analysis and compulsory schooling (detailed results available upon request). We find that lengthening

[6]Even if there are defiers of the described sort, we can at least argue that our estimates provide a lower bound estimate of the effect of education on health - assuming that causal effects of education on compliers and defiers have the same sign. Moreover, we can also reason that the latter group should be quantitatively unimportant compared to compliers, because they must come from the smaller group of intermediate track students.

compulsory schooling by one year increased the probability of attending middle school or academic track by around 1.5 percentage points for women and 0.6 percentage points for men. Thus, the net effect is positive but small, especially for men. While we cannot exclude the possibility that some students chose a less demanding school track as a reaction to increased compulsory schooling, we believe this is a minor concern. The main effect of the compulsory schooling reform appears to be the increase in the length of schooling for basic track students and not a big change in the student composition. Similarly, we created a dummy variable for leaving school without a degree and regressed this variable on all explanatory variables and compulsory schooling laws. The results are a bit more ambiguous. For men, we find that compulsory schooling laws decreased the likelihood of dropping out by around 0.25 percentage points (statistically significant on the 5% level), while for women we find a positive, albeit not significant effect of around 0.12 percentage points. Thus, compulsory schooling laws had not much of an effect on leaving school without certificate. If anything, fewer students leave without certificate. Again, this poses no problem for our analysis because in this case the monotonicity assumption is not violated.

1.6.4 Confounding effects of short school years

Another potential problem of our analysis is that the increase in compulsory schooling was accompanied by the introduction of short school years in some German states (see Pischke (2007) for a detailed description). In the years 1966-67 the start of the school year was shifted from Easter to the end of summer. All states except Bavaria, Hamburg, and West-Berlin introduced two short school years (of 8 months each) for students attending school at this time to accommodate the change in the schedule. Thus, although the first cohorts after the introduction of compulsory schooling have completed nine formal grades instead of eight, they have not spent much more time in the class room compared to the previous cohorts who were subject to eight (full) years of compulsory schooling. Bavaria did not introduce short school years because the school year already started in summer, and Hamburg and West-Berlin opted for one extra-long school year. The introduction of the short school years could affect our results because we possibly over-estimate the true extent of schooling of the affected cohorts. For this reason, we perform a robustness check by recoding our endogenous variable "years of schooling" taking into

Table 1.8: Robustness check with alternative definition of years of schooling taking into account short school years. Results of 2SLS regressions

Dependent variable		Sex	**First Stage coefficient on instrument**	**Second stage coefficient on years in school**	**N obs.**
Current health	Long-term illness	Men	0.666*** (0.088) *122.65*	-0.066*** (0.018)	19,824
		Women	0.579*** (0.096) *110.02*	0.006 (0.017)	20,080
	Work disability	Men	0.664*** (0.088) *120.19*	-0.054*** (0.015)	19,578
		Women	0.577*** (0.096) *107.94*	0.015 (0.016)	19,885
Weight	BMI	Men	0.591*** (0.079) *176.49*	-0.429*** (0.134)	43,509
		Women	0.622*** (0.075) *224.74*	-0.205 (0.175)	43728
	Overweight	Men	0.591*** (0.079) *176.49*	-0.046*** (0.018)	43,509
		Women	0.622*** (0.075) *224.74*	-0.033 (0.020)	43728
	Obesity	Men	0.591*** (0.079) *176.49*	-0.036** (0.014)	43,509
		Women	0.622*** (0.075) *224.74*	-0.004 (0.012)	43728
Smoking	Ever smoked	Men	0.593*** (0.076) *428.84*	-0.008 (0.013)	106,532
		Women	0.613*** (0.071) *548.26*	0.027** (0.013)	107,640
	Currently smoking	Men	0.595*** (0.076) *437.03*	-0.009 (0.013)	108,215
		Women	0.613*** (0.071) *563.55*	0.009 (0.013)	110,186
	Quitted smoking	Men	0.638*** (0.079) *353.71*	0.001 (0.013)	67,467
		Women	0.631*** (0.071) *268.39*	0.017 (0.016)	40,669

Note: Clustered standard errors (on cohort*state level) in parentheses. In addition, we report Kleibergen-Paap rK Wald F statistics in italics. All regressions include fixed effects for year of birth, state of residence, and an interaction of state of residence and a linear cohort trend. If several survey years have been pooled, fixed effects for survey year and the quadratic in age are included. Regressions with weight outcomes include height as additional control variable. * p<10 percent, ** p<5 percent, *** p<1 percent.

account actual time spent in school instead of highest grade completed. The results are shown in table 1.8.[7]

The results indicate that our estimates are not very sensitive to the exact definition of short school years. The result for smoking among women is an exception indicating that years of schooling increase the likelihood that women have ever smoked substantially. We can only speculate about the reason for this finding. Perhaps recoding the endogenous variable increased the weight of states without short school years in our estimates. Note, however, that this effect is transitory, as there is no statistically significant effect on current smoking among women.

1.7 Summary and discussion

The present paper contributes to the growing literature on identifying the causal link between education and health and health-related behavior. Economic theory has identified causal effects of education on health through several plausible channels: (a) education raises efficiency in health production; (b) education changes inputs into health production (through information) and thereby increases allocative efficiency; (c) education itself changes time preference (and thus inputs into health production) because schooling focuses students' attention on the future; (d) education has an indirect effect mediated through higher income, occupational status, better housing, or healthier environmental conditions.

Numerous studies have indeed documented a strong positive empirical association between education and health. Interpretation of this correlation as causal is difficult, however, because education is most likely an endogenous variable, for instances because unobserved "pre-treatment" variables such as time preferences or cognitive and social skills, possibly drive both education and health behavior, or because health at younger ages, e.g. in early childhood affects both educational achievement and later life health. Recent empirical work addresses causality issues head on using natural experiments such as exogenous changes in compulsory schooling laws for identification. Our paper is directly linked to this literature. Using data from several German Microcensuses, we exploit changes in years of compulsory schooling in West German federal states that took effect between 1949 and 1969 to estimate the causal effect of years

[7]Notice that the number of observations is lower in this sample because for individuals born in 1960 we cannot know for sure how many short school years they were exposed to.

in school on long-term illness, work disability, BMI (and overweight/obesity) and current and former smoking measured in 1989 to 2003.

We find evidence for a strong and significant negative causal effect of years of schooling on long-term illness and work disability among men. Our IV estimates are slightly larger than OLS but in contrast to some of the existing literature they tend to be in the same range. For women, however, we do not find any significant causal effects on health status. We also find some evidence for negative causal effects of education on male body weight and somewhat less convincing support for a negative effect on female body weight.

The literature on causal effects of education on health is usually silent on sex differences. However, there is a large literature on sex differences in health levels and their reasons (e.g. Verbrugge, 1989; Case and Paxson, 2005). Reasons for sex differences in health mentioned are (1) differences in biological risk, i.e. intrinsic differences based on human biology, (2) differences in acquired risk, i.e. health risks due to work and other activities, including health-related activities, (3) psychosocial differences, i.e. differences in health perception and believes, (4) differences in reporting styles, i.e. how illnesses and symptoms are communicated, and (5) differences in health care utilization and the effects of health care on current diseases and the onset of future diseases. One might hypothesize that education mediates those "pathways" differently for different sexes. For instance, education could mediate differences in acquired risk if women with low levels of education are less likely to perform hard physical labor than men. Thus, if education reduces the likelihood of being in a physically demanding job - and more so for men than for women - a stronger effect of education on health for men could result. Indeed, some supplementary analyses show that the effect of education on long-term health and work disability for men seems to be partly mediated through its effect on the probability of being a blue-collar worker. Overall, however, gender differences in the SES-health link appear to be underresearched. The five possible pathways mentioned in the preceding paragraph could serve as promising starting points.

We also address some possible concerns about the validity of our results. One concern is that, because we only know individuals' current state of residence and not the state where they actually finished school, migration between the states results in measurement error in both the instrument and schooling. Our robustness check - exploiting the fact that four large contiguous states have lengthened compulsory schooling in the same year - suggests that migration is unlikely to compromise our estimation results. Further, we address the concern of unobserved state-specific trends which could be correlated with the instrument are biasing our results. In our most flexible specification, we allow for (common) cohort fixed effects and cubic state-specific cohort trends. These robustness checks do not alter our main conclusions. Next, we address the concern that changes in mandatory schooling laws could also affect the likelihood that a student leaves school without a degree or induces track changes that actually reduce the number of years in school. Such behavior would violate the monotonicity assumption made in the presence of heterogeneous treatment effects. Regression results indicate very small effects on track choice that are unlikely to threaten the validity of our main results. Finally, we allow for the fact that students affected by the mandatory schooling reform also underwent so-called short school years. Recoding our endogenous variable to account for short school years does not alter our results either.

In conclusion, the results or our paper can be put in the context of the existing theoretical and empirical literature as follows: based on our analysis, we can only partly distinguish between the four theoretical arguments for a causal effect of education on health mentioned above. A link via better health inputs in terms of less smoking is not supported by our data.[8] Overweight and obesity can be interpreted not only as indicators of future health problems but also as indicators of past health behavior. Thus our findings on weight indirectly support the health input argument, although more so for men than for women. Again, changes in occupation from manual to non-manual could be an explanation. Further, we have not looked at income or wages in our paper. Considering existing evidence - using an identical identification strategy - changes in mandatory schooling had no causal effect on wages (Pischke and von Wachter (2008)), the link between education and health via higher income (and thus favorable living conditions) appears unlikely. Finally, it must also be noted that our parameters only identify the effect of education for compliers to the specific reforms of raising mandatory school leaving age. Interventions at other stages of the life-cycle or more specific interventions might have stronger and more systematic effects on health outcomes and health behavior.

[8] This is surprising because Jürges et al. (2011) do find an effect of education on smoking using similar data using the construction of academic track schools as instruments. Possibly, these differences arise because they look at a different part of the schooling distribution and because in their set-up the composition of peers in the schools also changes as more students are drawn into academic track schools. We leave this question to further research.

1.8 Appendix

Table 1.9: Reduced form results

Dependent variable		Sex	**Coefficient on reform dummy**	**Standard Error**	**N obs.**
Current health	Long-term illness	Men	-0.027**	0.011	23,128
		Women	0.006	0.010	23,479
	Work disability	Men	-0.021**	0.010	22,858
		Women	0.012	0.009	23,266
Weight	BMI	Men	-0.180**	0.073	48,640
		Women	-0.129	0.090	49,225
	Overweight	Men	-0.018**	0.009	48,640
		Women	-0.021**	0.010	49,225
	Obesity	Men	-0.018**	0.008	48,640
		Women	-0.003	0.007	49,225
Smoking	Ever smoked	Men	-0.007	0.007	119,461
		Women	0.006	0.007	121,541
	Currently smoking	Men	-0.003	0.010	121,318
		Women	-0.0003	0.006	124,314
	Quitted smoking	Men	-0.003	0.007	76,037
		Women	0.010	0.009	47,551

Note: Standard errors clustered on cohort*state level. * p<10 percent, ** p<5 percent, *** p<1 percent. All regressions include fixed effects for year of birth, state of residence, and an interaction of state of residence and a linear cohort trend. If several survey years have been pooled, fixed effects for survey year and the quadratic in age are included. Regressions with weight outcomes include height as additional control variable.

Chapter 2

Spillover effects of maternal education on child's health and health behavior[†]

2.1 Introduction

When analyzing returns to education, economists often focus on wages and income (see Card (1999) for an overview). More recently, research is also concentrating on the effect of education on non-market outcomes like health (see Cutler and Lleras-Muney (2008) and Grossman (2006) for overviews). Furthermore, researchers point to intergenerational spillover effects of education (Black and Devereux (2011) and Currie (2009) provide overviews). Quantifying such intergenerational links is not only relevant regarding optimal investments into education, but also relates to social mobility. The more that a child's outcomes are determined by its parents' education, the less that a society can be considered to be socially mobile.

Our paper investigates the effects of maternal education on child's health and health behavior in Germany. We consider both the effects on newborns and adolescents. Therefore, we look at various outcome variables: physical health, smoking behavior, overweight, and doing sports for adolescents; low birth weight and preterm birth for newborns. We apply an instrumental variables (IV) approach that has not yet been used in the intergenerational context. We instrument maternal education by the number of her siblings while conditioning on characteristics of her parents, the child's grandparents. For this purpose, we draw on a rich household survey, the German Socio-Economic

[†]This chapter is based on joint work with Jan Marcus, see Kemptner and Marcus (2013).

Panel Study (GSOEP), containing detailed information about *three* generations. We argue that, given the grandparents' characteristics, the number of the mother's siblings generates variation in maternal years of education that is exogenous regarding her child's health and health behavior. If grandparents are constrained in borrowing against the mother's future earnings, the number of her siblings affects household resources available for her educational investments.

Previous studies on the effects of parental education on child's health and health behavior in developed countries produced mixed evidence (see table 2.7 in the appendix). Currie and Moretti (2003) find maternal education reduces the risks of low birth weight and preterm birth. This finding is not corroborated by the IV-study of McCrary and Royer (2011). For teenage children, Carneiro et al. (2013) as well as Lindeboom et al. (2009) find no significant effects of parental education on the children's health status in their IV-analyses. Other studies for Germany analyze the intergenerational correlation of health (Coneus and Spiess, 2012), as well as the correlation between parental education and child health (Lamerz et al., 2005). We add to the literature by applying an IV strategy that works for the sample size of common household panels, by considering a variety of outcomes for both newborns and adolescents, and by investigating possible channels of the estimated effects. We focus only on mothers because the GSOEP basically reports on the partner of the mother and not on the biological father.

For newborns, we find maternal education to be associated with a reduced probability of preterm maturity. Our IV approach, however, does not indicate significant effects on newborns. For adolescents, we find strong and significant effects on health-related behavior for daughters. An additional year of maternal education is estimated to reduce the daughter's probability of smoking by 7.4 percentage points and to increase the daughter's likelihood of doing sports at least once a week by 7.5 percentage points. We do not obtain significant effects on sons' health behavior. We do not find any effects on child's physical health and overweight.

We demonstrate the robustness of our IV estimates by sequentially introducing the control variables. The results are not substantially altered when we include controls for grandparents' education, grandparent's occupational prestige and the size of the area where the mother grew up. Also the results do not change when we control for some possibly "bad controls" (i.e. variables that are possibly consequences of maternal education) like mother's fertility, health and health behavior. Furthermore, the results are robust to only considering mothers with one, two or three siblings as well as to more flexible specifications of the first stage. We discuss mother's health behavior, assortative mating, household income, and child's schooling track as possible channels of the estimated effects. Our results do not suggest that mother's health behavior, assortative mating or household income explain the effects on adolescent daughters. When including child's schooling track as an additional control variable, the effect of maternal education on daughter's smoking behavior disappears. Hence, maternal education seems to affect daughter's smoking behavior through the higher likelihood of the daughter pursuing a higher secondary schooling track. Even though early tracking is a peculiarity of the German schooling system, the mechanism at work (school quality or peer group effects) may also be relevant for other countries.

The paper is structured as follows. Section 2.2 describes the data and presents descriptive statistics. Section 2.3 contains a detailed discussion of our empirical strategy. In section 2.4, we present both Probit and IV- Probit results and present sensitivity checks. Section 2.5 investigates channels of the estimated effects. Section 2.6 concludes with a discussion on the implications of our findings.

2.2 Data and descriptive statistics

2.2.1 Sample

In our analysis we make use of the rich data from the German Socio-Economic Panel Study (GSOEP). The GSOEP started in 1984 and annually collects information at the household and individual levels (see Wagner et al., 2007). In 2010 more than 10,000 households participated in this panel study.

The GSOEP hosts several features that make it particularly attractive for the present analysis: Not only is it one of the largest and longest-running panel studies in the world, it also provides detailed health information on adolescents and newborns. Furthermore, due to the collection of additional biographical information of adult respondents, for children data on their parents and on their grandparents are available. We conduct our analysis for two different samples according to the child's age when the information was collected: "newborns" (0-18 months) and "adolescents" (18-19 years). Both samples are pooled across survey years. The following section describes these samples and the child outcomes in more detail, before turning to variables at the mother's and grandparents' level.

The sample of newborns is based on the "mother and child questionnaire", which the GSOEP introduced in 2003. It is distributed to the mothers of children born in the year of the survey or the year before. Therefore, children born from 2002 to 2010 constitute the newborns sample. The adolescents sample consists of West German children born between 1983 and 1992, using data from when these children were around 18 years of age. Hence, the sample is pooled across survey years. For health related variables, we use data from the year when the adolescents answered the relevant questions on the individual adult questionnaire for the first time. Since some of the health variables are only included every two years, for some adolescents we use information from the year they turned 18 and for the rest we use information from the year they turned 19. In the regressions we control for these age differences through fixed birth year effects.

2.2.2 Outcome variables

We look at six different health outcomes, two in the newborns sample, four in the adolescents sample. All outcomes are binary variables and coded in such a way that "1" reflects less healthy outcomes. In the newborns sample, we consider two different health indicators: preterm birth and low birth weight. Preterm birth refers to the birth of a child of less than 37 weeks gestational age. In developed countries, preterm birth is the major cause of infant mortality and neurological long-term morbidity (Martius et al., 1998). Another related health measure is the child's birth weight. Babies with low birth weight have adverse health status later on in life - even when controlling for preterm maturity (McIntire et al., 1999). We define a child to be of low birth weight if its weight at birth is below 3000g.[9] We only analyse biological children, and exclude twin babies from the newborns sample because their birth weight is lower in general (Naeye, 1964).

For adolescents, we construct a variable "overweight" indicating a body mass index (BMI) greater than 25. We code a binary variable "currently smoking" according to the question "Do you currently smoke, be it cigarettes, a pipe

[9] More commonly low birth weight is defined as a birth weight of less than 2500. However, we encounter the same problem as Lindeboom et al. (2009, 111): with this strict definition we only have a few observations with low birth weight. These observations might be affected by measurement error. Hence, we apply the same definition as Lindeboom and colleagues.

or cigars?" The GSOEP started asking detailed health questions, including weight and smoking behavior, in even numbered years, starting in 2002.[10] A variable on sport activities indicates whether an adolescent is *not* doing sports at least once a week.[11] We generally use information on sport behavior from the year the adolescent turned 18. However, this variable was not collected in the survey years 2002, 2004, 2006, and 2010. For those who turn 18 in these years, we use the information about doing sports from the year they turned 19 - information gathered during the next wave. Apart from these three variables indicating health behaviors, we also look at a measure of general health status for adolescents. Our measure is based on the physical component summary scale (pcs) provided by the GSOEP group, a weighted combination of the 12 items of the SF-12 module (Andersen et al., 2007). In order to facilitate comparison to the other outcomes we also dichotomize this outcome variable. Adolescents with physical scale values below the median of all adolescents are coded as having "poorer health".[12]

2.2.3 Parental and grandparental variables

At the parental level, we focus only on mothers because the GSOEP collects data on the mother's partner, who may or may not be the biological father of the child. Relevant data for mothers include years of education, number of siblings and population of the area where the mother grew up until the age of 15. The GSOEP constructs the years of education variable from the respondents' information about the obtained level of education and adds time for additional occupational training.[13] For the numbers of siblings, we use

[10]The GSOEP collects data on smoking behavior also in the years 1998, 1999 and 2001. These questions, though, differ in their phrasing. Therefore, we exclude information from these survey years.

[11]We also computed the regressions for a slightly different definition of this variable. The results differ only marginally, when we consider a person as being active who is doing sports at least once a *month*. These and other results not shown are available from the authors upon request.

[12]We also applied different thresholds for the definition of poorer health, used the metric instead of the dichotomized physical health measure and resorted to a different self-rated health item ("How would you describe your current health?"). All these redefinitions of the health status outcome do not change the results presented in section 2.4.

[13]If the variable years of education is missing for an individual in a given survey year, we use information from other survey years. Following Kemptner et al. (2011), we also employ a different measure of the years of education, in which we only considered years of primary and secondary schooling: 9 years for individuals without school degree and those with basic track degree, 10 years for those with intermediate track school or other degree, 12 years for technical school degree and 13 years for general university-entrance diploma. However, the results did not change qualitatively, only the size of the coefficient estimates increased.

the earliest available information about brothers and sisters collected in the survey.[14] Since siblings might have died, this is the best approximation of the number of brothers and sisters when the mother went to school. The area where the mother grew up is a discrete variable with four categories according to the size of the hometown: countryside, small city, medium city and large city. All information about mothers is self-reported by the mothers.

At the level of the parents of the mother, we use data on educational levels and occupational prestige. For both, grandfathers and grandmothers, we construct dummy variables according to five educational levels: secondary school degree, intermediate/technical school degree, general university-entrance diploma, other degree and no school degree/no school attended. To measure occupational prestige we rely on the International Socio-Economic Index of Occupational Status (ISEI). The ISEI assigns scores to almost 300 different occupation categories "in such a way as to maximize the role of occupation as an intervening variable between education and income" (Ganzeboom et al., 1992, 2).[15] Information on the grandparents is either contributed by the grandparents directly (less than 5 percent) - if they are GSOEP participants - or by proxy via interviews of the mothers: All individuals with a valid personal interview in the GSOEP are requested to answer the supplementary biography questionnaire with questions on their parents and their social origin. Missing values at the grandparental level are imputed as described in the appendix.

2.2.4 Descriptive statistics

Table 2.1 displays unweighted means and standard deviations for relevant variables at the maternal and child level for both of our samples. While the newborns sample consists of West German children born between 2002 and 2010 and excludes both adopted children and twins, the adolescents sample consists of West German children born between 1983 and 1992. Both samples do not include children whose mothers were educated in the German Democratic Republic. We make the estimation samples more homogeneous by restricting them to mothers with siblings (see subsection 2.3.2 for further discussion). Due

[14] The GSOEP collected this information in 1990, 1996, 2001, 2003 and 2006. We consider siblings inside and outside of the household.

[15] The ISEI score is derived from the occupational status of grandfather and grandmother. GSOEP questions on the occupational status of grandfather and grandmother are formulated to reflect the situation when the mother was 16. For each pair of grandparents we make use of the highest ISEI score, which in most cases is the score of the grandfather. Missing values are imputed as described in the appendix.

Table 2.1: Descriptive statistics

Variable	Adolescents sample		Newborns sample	
	Mean	SD	Mean	SD
Mothers				
Years of education	11.9	2.7	12.8	2.8
Year of birth	1959.9	5.4	1974.5	5.9
Number of siblings	2.6	1.9	2.1	1.7
Children				
Year of birth	1987.5	2.8	2005.1	2.3
Birth weight			3353.4	558.3
Preterm birth			16.2	36.9
Low birth weight			19.5	39.7
Currently smoking (%)	27.6	44.7		
Overweight (%)	17.8	38.2		
No sport (%)	44.8	49.7		
Poorer health (%)	47.4	49.9		
N	1741		977	

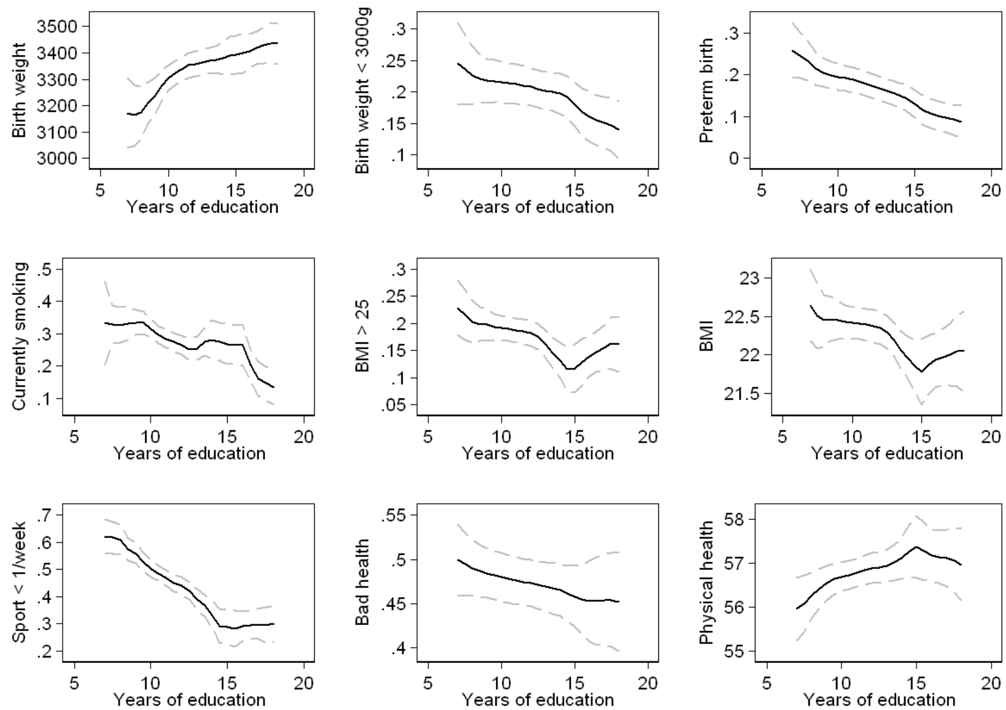
Note: Unweighted means and standard deviations for key variables of the sample of 18/19 year olds (adolescents) and of 0-18 month olds (newborns) as well as their mothers.

to the construction of the two samples, the mothers of the adolescents come from earlier birth cohorts. Therefore, differences in mean years of education and number of siblings between the two samples can be explained by the increase in years of education and the decrease of family sizes over time. All our regression models include the mother's year of birth to control for these time trends.

Figure 2.1 presents lines from non-parametric local constant estimations of the association between mother's years of education and various child outcomes.[16] For almost all outcome measures there is a monotonous relationship: Worse health behavior and poorer health of the child decrease almost linearly with the mother's education. For instance, the chance of not doing sports at least once a week is around 60 % for children of poorly educated mothers, 50 % for children of mothers with about 10 years of education and 30 % for children of mothers with more than 15 years of education. The probability of preterm birth is almost twice as high for the least educated mothers compared

[16]The local constant estimators rely on a plugin estimator of the asymptotically optimal constant bandwidth (see Fan and Gijbels, 1996; StataCorp, 2009) and an Epanechnikov kernel

Figure 2.1: Bivariate relationships between maternal education and child's outcomes



Note: The lines picture regression lines from non-parametric local constant estimators as well as the 95% confidence bands.

to the best educated mothers. The increase in the overweight probability on the right tale of the education distribution is not statistically significant as the confidence bands indicate.

2.3 Empirical strategy

2.3.1 Main estimation equation

We estimate the effects of maternal years of education on binary child outcomes. For this purpose, we rely on single (Probit) and two-equation models (IV-Probit).[17] All models are estimated with robust standard errors that are clustered by mothers, accounting for serial correlation between children of the same mother. Our single equation model linking child's outcome to maternal

[17]In the section on robustness checks, we also present results from a two-stage least squares model (2SLS). Being more robust regarding the distributional assumptions of the error terms but less efficient, the estimated effects differ only marginally.

years of education is specified as follows:

$$H_c = 1\left[\beta_0 + \beta_1 \cdot S_m + \boldsymbol{\beta_2} \cdot \mathbf{x} + \varepsilon_c > 0\right] \tag{2.1}$$

where H_c is child's outcome and S_m is maternal years of education. $\mathbf{x}$ contains different sets of control variables that we gradually incorporate in section 2.4 to assess their impact. In the preferred specification, $\mathbf{x}$ includes sets of fixed effects accounting for mother's birth cohort, the size of the area where the mother grew up, grandparents' level of education and occupational prestige, child's birth cohort, and child's sex.[18] ε_c is an idiosyncratic child specific error term that is normally and identically distributed. 1[ů] is an indicator function.

Estimating equation (2.1) as a single equation model will only produce consistent parameter estimates if maternal years of education, S_m, are uncorrelated with ε_c. Since maternal years of education are likely to be correlated with unobserved confounders, we expect the coefficient estimates to be biased in an unknown direction.

2.3.2 Instrumental variables approach

The endogeneity of S_m can be dealt with by instrumenting S_m with Z_m, where Z_m must meet the following two conditions:

$$E[\varepsilon_c|Z_m] = 0 \qquad \textbf{(validity)}$$
$$E[S_m|Z_m, \mathbf{x}] \neq E[S_m|\mathbf{x}] \qquad \textbf{(relevance)}$$

Z_m is a valid instrument if it affects the child's outcome only through mother's years of education, given the other covariates. Z_m is a relevant instrument if the explanatory power of Z_m with respect to S_m is sufficiently large, given the other covariates. Various instruments for education are proposed in the literature on returns to education (see Card (1999) and Grossman (2006) for overviews). A first wave of IV studies relies on family characteristics as instruments, such as parents' income and parents' schooling. While these instruments are strongly associated with education, the validity assumption seems questionable. A second wave of IV studies uses variations in educational policies and other natural experiments. This second wave of IV studies faces

[18] By controlling for both mother's birth cohort and child's year of birth, we indirectly control for the mother's age at birth. Since the mother's age at child birth is a choice variable and possibly correlated with maternal education, we also run the models without this control variable. The results are insensitive to this modification. These estimates are available upon request from the authors.

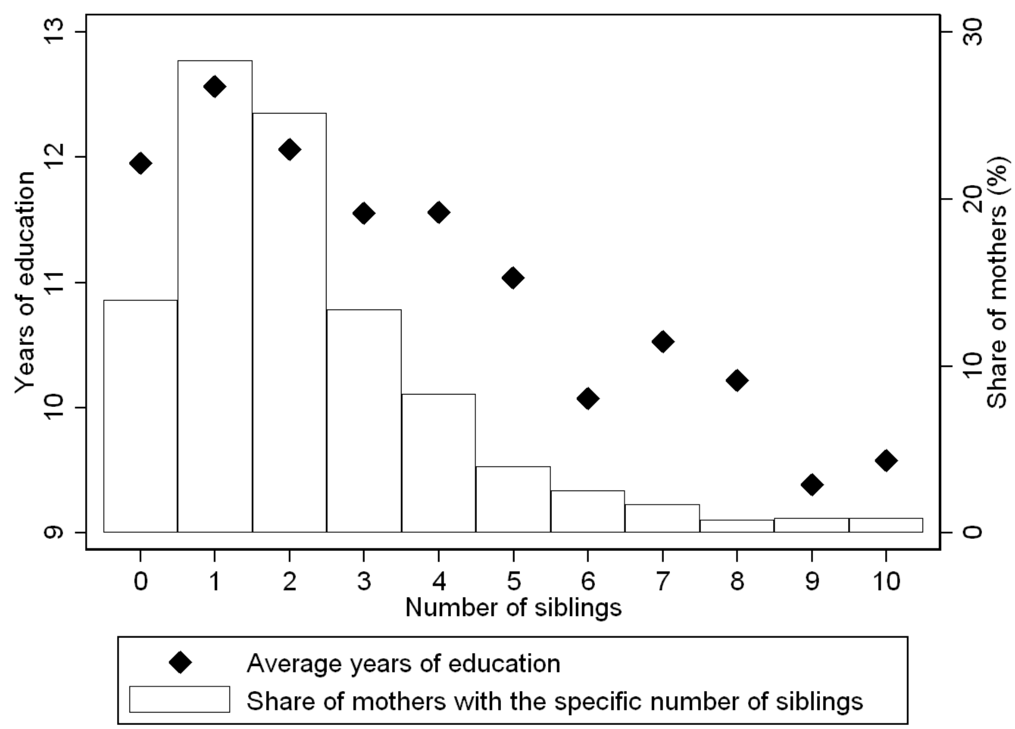
less criticism regarding the validity assumption. However, the association with education is often weak and, hence, weak instrument problems may arise. Researchers frequently draw on huge sample sizes to mitigate this problem. A drawback of huge data sets is that these often do not include detailed outcome measures. Another problem with policy changes and other natural experiments is that they only affect certain cohorts.

We do not rely on policy changes but instead use the number of mother's siblings as an instrument for maternal education while conditioning on characteristics of the grandparents. These grandparental characteristics include variables describing the grandparents' level of education and occupational prestige as well as the area where the mother grew up. This identification strategy works also for cohorts unaffected by policy changes and for the limited sample sizes of common household panels. This instrument was suggested before (e.g. Sander, 1995). We improve the approach by conditioning on characteristics of the grandparents. There is an obvious concern regarding the validity of the instrument. Fertility is higher in the countryside and negatively correlated with social status, i.e. mothers with siblings are more likely to live in the countryside and to have parents with lower social status. Therefore, we condition on the grandparent's level of education, the grandparent's occupational prestige and the size of the area where the mother grew up. We gradually incorporate these variables to assert their influence.

Consistency of our estimates rests on the assumption that the instrument identifies exogenous variation in the endogenous education variable, given the other covariates. We deal with possible concerns regarding the validity of the instrument by including controls for maternal health, health behavior and fertility in one specification. However, this is not our preferred specification because the additional control variables might be consequences of maternal education themselves.

The number of mother's siblings should also be a relevant instrument because the resources available for educational investments per child depend substantially on the number of children in the household. This assumes that parents are constrained in borrowing against their children's future earnings. A significant effect of the number of mother's siblings on her education in the first stage points to such a borrowing constraint of the grandparents. Even though there are no schooling fees and very low or no tuition fees at public educational institutions in Germany, investments into children's education involve forgone earnings for both the parents and the children. Parents' time constraints and limited housing space may impose pressure upon the children to make their own living instead of spending more time on educational investments.

Figure 2.2: Number of siblings and years of education



Note: Average years of education for mothers with different numbers of siblings as well as the share of mothers with that number of siblings.

Figure 2.2 contains a graph showing the average years of education for mothers with different numbers of siblings as well as the share of mothers with this number of siblings.[19] The graph shows that most children in our sample are born to mothers who grew up with four or fewer siblings (about 85 %). Furthermore, the graph pictures a clear negative relationship between maternal education and the number of her siblings. However, mothers without siblings seem to be special having on average less education than mothers with one or two siblings. This does not correspond with the argument that more siblings introduce resource constraints leading to lower educational achievements. Black et al. (2005) also find this only child particularity for the US, which disappears when they consider the subsample of intact families. Hence, it may be that parental divorces exert negative effects on the only children's educational achievement. The only child particularity could lead to a non-linear relation-

[19]The graph displays the numbers for the sample of adolescents with non-missing smoking information. Similar pictures emerge for the other samples.

ship between mother's years of education and the number of her siblings or even have long-lasting effects on the grandchildren. For this reason, and in order to make the estimation sample more homogeneous, we restrict the sample to mothers with siblings. In subsection 2.4.3, we present a robustness check for the inclusion of children of mothers without siblings. It turns out that our findings are insensitive to this modification of the sample.

2.3.3 Implementation

We implement the IV strategy by estimating the following two-equation model using the method of maximum likelihood:

$$S_m = \gamma_0 + \gamma_1 \cdot Z_m + \boldsymbol{\gamma_2} \cdot \mathbf{x} + \mu_m \tag{2.2}$$

$$H_c = 1\left[\beta_0 + \beta_1 \cdot S_m + \boldsymbol{\beta_2} \cdot \mathbf{x} + \varepsilon_c > 0\right] \tag{2.3}$$

μ_m and ε_c are assumed to be bivariate normally and identically distributed with mean zero and to be independent of the instrument Z_m (see Wooldridge (2002, 472) for further details). Under the assumptions of instrument validity and relevance, joint estimation of equations (2.2) and (2.3) as an IV-Probit model produces consistent parameter estimates. The coefficients of the first stage can be directly interpreted as marginal effects. Since the parameters of a Probit model cannot be given this interpretation, we compute average marginal effects and apply the delta method when calculating standard errors. In subsection 2.4.3, we present a robustness check for the distributional assumptions of the model.

2.4 Results

Table 2.2 contains the first stage results of the IV-Probit model. The second stage results of the IV-Probit model and the findings from the single equation Probit model are presented in table 2.3 for the newborns and table 2.4 for the adolescents. As has been discussed above, we estimate several specifications and sequentially introduce the control variables. This demonstrates the robustness of our estimates. Note that for the IV-Probit model specification 6 is our preferred specification because it conditions on a rich set of grandparental characteristics, but does not include potentially "bad controls" like mother's health, health behavior, and fertility. In the following subsections, we first discuss the effect of mother's siblings on her educational achievement (first stage of the

IV-Probit model), then we discuss the findings on the effects of maternal education on the child's outcomes (Probit model and second stage of the IV-Probit model), and subsequently we present some additional sensitivity checks.

2.4.1 The effect of siblings on years of education

Table 2.2 presents the first stage coefficients of the IV estimation. The small differences in the first stage coefficient estimates stem from different sample sizes for the outcome measures. The estimated effects of the number of mother's siblings on her educational attainment are highly significant in all specifications. This indicates that the number of siblings is a relevant instrument for maternal education. The association is slightly stronger for adolescent daughters than for adolescent sons, which is presumably due to sampling variation.[20] Conditioning on characteristics of the grandparents' household reduces the gradient between the number of siblings and educational achievement of the mothers. In the preferred specification 6, an additional sibling is predicted to decrease the years of education by 0.23-0.28 years for the female adolescents and the newborns samples and by about 0.15 years for the male adolescents sample. We are interested in the F-statistics testing the assumption that the number of mother's siblings does not affect her educational achievement, given the other covariates. For specification 6, all our F-statistics for the pooled newborns sample are above 24. Furthermore, all the F-statistics are above 37 for the sample of female adolescents and above 11 for the sample of male adolescents.

Thus, the estimation strategy seems not to suffer from a weak instruments problem. The estimated significant effects in the first stage point to financial constraints of the grandparents when investing in their daughter's education.

2.4.2 The effect of maternal education on child outcomes

Newborns

In the Probit models, the average marginal effects indicate a significant association between maternal education and the likelihood of preterm birth (see specification 1-3 in table 2.3). When conditioning on grandparental characteristics, we find no evidence for an effect of maternal education on low birth weight.[21] Note that the grandparental characteristics capture some potential

[20]None of the effect differences between mothers of daughters and mothers of sons is significant at the 5 % level.

[21]We also obtain small and insignificant effects when using birth weight as outcome in 2SLS regressions.

Table 2.2: First stage - the effect of number of siblings on years of education

Sample	Obs.	(4)	(5)	(6)
Newborns				
(Preterm birth)	962	−0.413***	−0.316***	−0.282***
		(0.069)	(0.057)	(0.057)
		[35.77]	[31.07]	[24.58]
(Low birth weight)	977	−0.402***	−0.311***	−0.280***
		(0.067)	(0.055)	(0.055)
		[36.06]	[32.10]	[25.72]
Adolescent daughters				
(Smoker)	867	−0.391***	−0.253***	−0.237***
		(0.041)	(0.038)	(0.037)
		[92.76]	[45.34]	[40.70]
(Overweight)	859	−0.390***	−0.250***	−0.233***
		(0.042)	(0.039)	(0.038)
		[86.68]	[41.82]	[37.39]
(No sport)	793	−0.378***	−0.260***	−0.249***
		(0.042)	(0.039)	(0.039)
		[81.18]	[45.21]	[41.69]
(Poorer health)	843	−0.393***	−0.249***	−0.234***
		(0.042)	(0.038)	(0.038)
		[88.87]	[42.76]	[38.85]
Adolescent sons				
(Smoker)	874	−0.297***	−0.170***	−0.149***
		(0.041)	(0.039)	(0.039)
		[51.73]	[18.91]	[14.62]
(Overweight)	851	−0.291***	−0.158***	−0.137***
		(0.041)	(0.039)	(0.039)
		[49.32]	[16.19]	[12.34]
(No sport)	805	−0.349***	−0.211***	−0.182***
		(0.045)	(0.042)	(0.042)
		[60.77]	[24.59]	[18.60]
(Poorer health)	836	−0.291***	−0.159***	−0.136***
		(0.042)	(0.040)	(0.040)
		[48.46]	[15.93]	[11.77]
GP education			Y	Y
GP status				Y

Note: First stage results. Marginal effects of the number of mother's siblings on her years of education, robust standard errors (in parentheses) and F-statistics (in brackets) separately for mothers of newborns, adolescent daughters and adolescent sons. All regressions include controls for the child's and the mother's year of birth. Specifications 2 and 5 include additional fixed effects for the area the mother grew up and for the educational levels of the mother's parents, while specifications 3 and 6 control additionally for the grandparents' occupational prestige (ISEI score). * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

confounders of maternal education being related to the family background. The average marginal effect on preterm birth matches the smoothed bivariate regression line from figure 2.1. Controlling for characteristics of the grandparents' household, an additional year of maternal education is associated with a reduction in the probability of preterm maturity by 1.3 percentage points (specification 3).

Although insignificant for most specifications, in the IV model the estimated effect on preterm maturity increases in size compared to the findings of the single equation Probit model (2.3 vs. 1.3 percentage points reduction in risk). It might be that our sample size is too small to detect an existing effect. For the US, Currie and Moretti (2003) report a significant reduction of 1 percentage point in the probability of preterm birth for each year of maternal education (see table 2.7 for an overview of the results and the designs of previous IV-studies on the effect of maternal education on child's health in developed countries).

Average marginal effects on low birth weight are close to zero and the standard errors are relatively large. This finding is in line with Lindeboom et al. (2009). Making use of a change in compulsory schooling in Britain in 1947, they find no effect of mother's education on birth weight and low birth weight. With the same policy change, Chevalier and O'Sullivan (2007) estimate an increase of 74 gram birth weight for every additional year of mother's education but do not find a significant reduction in the probability of low birth weight. Also McCrary and Royer (2011) and Carneiro et al. (2013) find no effect on low birth weight. Only Currie and Moretti (2003) estimate a significant reduction of 1 percentage point in the probability of low birth weight for an additional year of maternal education.

Adolescents

The Probit specifications in table 2.4 indicate that one more year of mother's education is associated with a decrease in the adolescent's probability of being a smoker by about 2 percentage points, given the grandparents' characteristics. There is no significant association between years of maternal education and an adolescent's probability of being overweight at age 18/19. However, there seems to be a strong relationship with the child's likelihood of not doing sports at least once per week. The estimates suggest that each additional year of maternal education is associated with a decrease in the probability of not doing sports regularly by 3 percentage points for sons and 3.4 percentage points for daughters. The results do not suggest an association of maternal education

Table 2.3: The effect of maternal education on newborns' outcomes

		Probit			IV-Probit		
Outcome	Obs.	(1)	(2)	(3)	(4)	(5)	(6)
Preterm birth	962	−0.020***	−0.014 **	−0.013 **	−0.037 **	−0.022	−0.023
		(0.005)	(0.006)	(0.006)	(0.016)	(0.022)	(0.025)
Low birth weight	977	−0.011 **	−0.007	−0.005	0.005	0.014	0.021
		(0.005)	(0.006)	(0.006)	(0.020)	(0.025)	(0.028)
GP education			*Y*	*Y*		*Y*	*Y*
GP status				*Y*			*Y*

Note: Average marginal effects of maternal education and robust standard errors (in parentheses) for Probit (1-3) and second stage of IV-Probit (4-6) models for the sample of newborns. All regressions include controls for the child's and the mother's year of birth as well as for the child's sex. Specifications 2 and 5 include additional fixed effects for the area the mother grew up and for the educational levels of the mother's parents, while specifications 3 and 6 control additionally for the grandparents' occupational prestige (ISEI score). * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

with an adolescent's physical health.

Turning to the results of the IV-Probit model (specifications 4-6 in table 2.4), we find large and significant effects of maternal years of education on daughter's smoking and sport behavior in all IV specifications. These effects do not disappear when we include the grandparents' level of education, their occupational prestige and the size of the area the mother grew up as additional controls. The coefficients tend to be even slightly larger. In our preferred specification 6, the probability of the daughter doing sports regularly is increased by 7.4 percentage points per year of maternal education. In addition, one additional year of maternal education decreases the likelihood of the daughter being a smoker at age 18/19 by 7.5 percentage points. We do not find any significant effects on son's health and health behavior in general.[22] Confirming the finding from the Probit specification, there is also no significant effect on overweight or physical health for girls. Loureiro et al. (2010) also find that mothers are only influential with respect to the smoking behavior of their daughters but not for their sons. These gender differences are in line with the idea of gender-specific parental role-models and the finding that children identify stronger with the same-sex parent (Starrels, 1994). Furthermore, this supports the theory of more productive parenting by the same-sex parent (Gugl and Welling, 2011).

Comparing the significant effects of maternal education on child's outcomes with the estimates from the Probit models, the estimated effects from the IV-Probit models appear to be larger. This is in line with the majority of findings in the literature on returns to education (see Card, 1999). Currie and Moretti (2003) and Carneiro et al. (2013) also find larger effects when instrumenting maternal education. Three factors might be responsible for this finding. First, measurement error in maternal education attenuates the Probit estimates. Second, unobserved variables that are negatively correlated with maternal education but positively with better child outcomes might result in downward biased estimates. Third, in the presence of effect heterogeneity, IV approaches may not identify the average effect for the overall population but rather local average effects for the so-called compliers, i.e. mothers who obtain fewer [more] years of education because they have more [less] siblings.

[22] Although the effect on sport activity is rather large and only borderline insignificant.

Table 2.4: The effect of maternal education on adolescents' outcomes

		Probit			IV-Probit		
Outcome	Obs.	(1)	(2)	(3)	(4)	(5)	(6)
Daughters							
Smoker	867	−0.017***	−0.020***	−0.018 **	−0.046***	−0.074***	−0.075***
		(0.006)	(0.007)	(0.007)	(0.016)	(0.022)	(0.024)
Overweight	859	−0.008*	−0.009	−0.006	0.017	0.029	0.031
		(0.005)	(0.006)	(0.006)	(0.017)	(0.027)	(0.029)
No sport	793	−0.040***	−0.038***	−0.034***	−0.072***	−0.076***	−0.074***
		(0.006)	(0.008)	(0.008)	(0.016)	(0.026)	(0.028)
Poorer health	843	−0.010	−0.008	−0.006	−0.009	−0.010	−0.009
		(0.007)	(0.008)	(0.008)	(0.023)	(0.039)	(0.042)
Sons							
Smoker	874	−0.015 **	−0.018***	−0.019***	0.002	0.013	0.015
		(0.006)	(0.007)	(0.007)	(0.027)	(0.049)	(0.056)
Overweight	851	−0.008	−0.005	−0.004	−0.020	−0.020	−0.010
		(0.006)	(0.007)	(0.007)	(0.025)	(0.048)	(0.057)
No sport	805	−0.026***	−0.030***	−0.030***	−0.031	−0.052	−0.056
		(0.007)	(0.008)	(0.008)	(0.024)	(0.038)	(0.043)
Poorer health	836	0.000	0.001	0.001	0.005	0.021	0.024
		(0.007)	(0.008)	(0.008)	(0.031)	(0.058)	(0.067)
GP education			*Y*	*Y*		*Y*	*Y*
GP status				*Y*			*Y*

Note: Average marginal effects of maternal education and robust standard errors (in parentheses) for Probit (1-3) and second stage of IV-Probit (4-6) models separately for mothers of boys/girls. All regressions include controls for the child's and the mother's year of birth. Specifications 2 and 5 include additional fixed effects for the area the mother grew up and for the educational levels of the mother's parents, while specifications 3 and 6 control additionally for the grandparents' occupational prestige (ISEI score). * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

2.4.3 Sensitivity checks

Table 2.5 presents five sensitivity checks for the adolescents samples.[23] These additional specifications are estimated to show that some possible concerns regarding the instrument's validity, the sample characteristics, the functional form assumptions, or the distributional assumptions of the error terms are unlikely to compromise our results.

Specification 7 includes additional controls' for mother's health, health behavior and fertility. Since mother's health and health behavior are potentially affected by the number of her siblings, we include the four health-related outcome variables - at the mother's level - as additional control variables, i.e. we include binary variables for the mother's overweight, smoking, sport behavior and bad health status. A further concern for the instrument's validity relates to mother's fertility. Grandparents' fertility could affect mother's fertility and lead to financial constraints that have a direct impact on child outcomes. Indeed, we find that the number of mother's siblings explains 4 % of the variation in the number of mother's own children. We address this concern by including fixed effects for the number of mother's children in our IV model.[24] However, this is not our preferred specification as the additional control variables might be inherent consequences of maternal education and, hence, bad control variables. The estimated effects of specification 7 differ only marginally from the estimates of our preferred specification 6.

Specification 8 allows for full flexibility with respect to the functional relationship between the number of siblings and maternal education. In this specification we instrument maternal education by a set of dichotomous variables that indicates the number of siblings. Categories of the number of siblings are 1, 2, 3, 4, 5 and 6 and more siblings. Specification 9 includes children of mothers without siblings to show that our findings do not hinge on this sample restriction.

Another concern regarding our instrument is that the fertility decision of parents may be affected by heterogeneous preferences for child quality if parents take into account the resource constraints of the household ("quantity-quality trade-off"; see Becker et al., 1960; Becker and Lewis, 1973). In principle, conditioning on the grandparental characteristics should account for this heterogeneity to the degree that the preferences for child quality are correlated with

[23]We do not show sensitivity analyses for newborns because similar to the results in table 2.3 the effects are insignificant in all sensitivity analyses.

[24]More specifically, we include dummy variables for 1, 2, 3 and 4 or more children.

Table 2.5: The effect of maternal education on adolescents' outcomes - robustness & channels

	Robustness					Channels		
Outcome	[add. controls] (7)	[dummies] (8)	[+only child] (9)	[1-3 siblings] (10)	[2SLS] (11)	[mating] (12)	[income] (13)	[ch. educ.] (14)
Daughters								
Smoker	−0.070 **	−0.061 **	−0.093***	−0.110***	−0.062*	−0.064	−0.083***	−0.019
	(0.029)	(0.030)	(0.023)	(0.035)	(0.034)	(0.040)	(0.030)	(0.040)
Overweight	0.041	−0.002	0.059*	0.049	−0.001	0.065	0.052	0.046
	(0.034)	(0.032)	(0.031)	(0.068)	(0.028)	(0.042)	(0.035)	(0.037)
No sport	−0.083***	−0.073 **	−0.070 **	−0.085*	−0.071*	−0.072	−0.074 **	−0.077 **
	(0.029)	(0.036)	(0.035)	(0.048)	(0.037)	(0.048)	(0.037)	(0.032)
Poorer health	−0.017	−0.015	−0.017	−0.004	−0.016	−0.044	−0.009	−0.009
	(0.049)	(0.039)	(0.050)	(0.093)	(0.040)	(0.055)	(0.054)	(0.053)
Sons								
Smoker	0.017	0.031	−0.057	0.025	0.034	0.066	0.037	0.000
	(0.081)	(0.049)	(0.052)	(0.094)	(0.050)	(0.107)	(0.067)	(0.075)
Overweight	−0.001	−0.007	−0.065	0.090 **	−0.010	0.002	−0.000	−0.033
	(0.072)	(0.057)	(0.048)	(0.041)	(0.046)	(0.113)	(0.078)	(0.073)
No sport	−0.020	−0.092***	−0.010	0.045	−0.101 **	−0.050	−0.052	−0.058
	(0.079)	(0.032)	(0.060)	(0.097)	(0.050)	(0.128)	(0.059)	(0.058)
Poorer health	0.039	−0.008	0.007	−0.052	−0.006	0.015	0.026	0.022
	(0.081)	(0.061)	(0.076)	(0.084)	(0.053)	(0.140)	(0.093)	(0.091)

Note: See specification 6 in table 2.4. Specification 7 includes additional controls for maternal health, health behavior and fertility. Specification 8 rests upon dummies for 1, 2, 3, 4, 5 and 6+ siblings as instruments. Specification 9 includes also mothers without siblings, while specification 10 only considers mothers with 1, 2 or 3 siblings. Specification 11 is estimated with two-stage least squares. Specification 12 controls for the years of education of the mother's partner, specification 13 includes the logarithm of a five years average of household income and specification 14 controls for whether the child is on an academic track school. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

education or occupational prestige. The concern, however, is that conditioning on these variables might not be enough. In specification 10, we estimate our model for a more homogeneous sample that includes only mothers with one to three siblings. This sensitivity check relies on the assumption that parents of these mothers are presumably more similar than in the full sample. We find it reassuring that our results are not sensitive to this sample restriction.

Lastly, we check the sensitivity of our findings regarding the distributional assumptions of the IV-Probit model (specification 11). We estimate a two-stage least squares model (2SLS), which - unlike the IV-Probit model - also produces consistent parameter estimates in the presence of heteroscedasticity and non-normally distributed errors.

In all these additional specifications, most of the estimated effects change only marginally. The effects of maternal education on sport and smoking behavior of daughters are significant in all sensitivity tests. The effect of maternal education on sons not doing sports regularly becomes significant for specification 8 and specification 11. When only considering mothers with one, two or three children, a positive effect of mothers on the son's probability of being overweight emerges (specification 10). This is likely to be due to a weak instruments problem that arises for sons when using only the restricted sample. The first stage F-statistics for this specification is about 5 for sons, and about 10 for daughters. We conclude that our main findings, the effects on daughters' smoking and sport behavior, are not sensitive regarding sample characteristics, functional form assumptions or distributional assumptions of the error terms.

Another issue is that there may be non-linearities in the effects of maternal years of education on child's outcomes, although the graphs in figure 2.1 suggest an approximately linear relationship. We try to detect non-linearities in the effects by using years of education and years of education squared as endogenous regressors in our IV-models. Accordingly, we instrumented the two endogenous regressors with the number of mother's siblings and its square. The estimated coefficients on the years of education squared were highly insignificant. Non-linearities in the effects of maternal years of education are either irrelevant or too small to be detected with our estimation approach and the given sample size. In any case, our estimates can be interpreted as the *average* effect of one more year of maternal education.

2.5 Channels

This section discusses possible channels of the estimated effects that could drive the relationship between maternal education and adolescents' outcomes. In order to investigate potential channels, we follow the strategy by Oreopoulos et al. (2008). When investigating intergenerational effects of father's displacement, Oreopoulos et al. (2008) analyze potential channels by a) including the channels as (potentially endogenous) additional controls and by b) investigating the effect on the channels (as outcomes).

We consider mother's health behavior, assortative mating, household income, and child's schooling track as potential channels. Table 2.5 contains the results of alternative specifications that include these potential channels as additional controls. The results from these specifications must be interpreted carefully because these additional controls are likely to be endogenous and may also bias the estimated effects of maternal education. Furthermore, table 2.6 presents estimates of the direct effects of maternal education on the potential channels using our IV approach.

Mother's health behavior may explain the estimated effects of maternal education on daughter's health behavior if the mother operates as role model. We look at the same health measures that we also consider for adolescents. Indeed, we find substantial effects of maternal years of education on mother's own health behavior, but not on physical health, using our IV approach (see table 2.6).[25] However, the effects of maternal education on daughter's health behavior remain unchanged when controlling for mother's health behavior (smoking, overweight, and no sports) in the IV model (see table 2.5, specification 7).

Assortative mating may explain to some degree the effects of maternal education on child outcomes. In our data we find a correlation coefficient of 0.66 between maternal years of education and her partner's years of education. Furthermore, our IV approach predicts that one more year of maternal education increases partner's years of education by 0.84 years (table 2.6). Thus, the estimated effects on daughter's health behavior may work through the partner's education. We focus on the mother's partner because the GSOEP does not report on the biological father, just on the mother's current partner. In specification 12 (table 2.5), we estimate effects of maternal education on the child's outcomes, including the partner's years of education as an additional control

[25] We do not find evidence that these effects differ substantially between mothers of daughters and mothers of sons.

Table 2.6: The effect of maternal education on potential channels

		Probit/OLS			IV-Probit/2SLS		
Outcome	Obs.	(1)	(2)	(3)	(4)	(5)	(6)
Mothers' health							
Mother smokes	1794	−0.029***	−0.036***	−0.036***	−0.021	−0.051 **	−0.054*
		(0.004)	(0.005)	(0.005)	(0.016)	(0.025)	(0.028)
Mother overweight	1798	−0.027***	−0.021***	−0.021***	−0.077***	−0.088***	−0.095***
		(0.005)	(0.005)	(0.005)	(0.009)	(0.016)	(0.016)
Mother no sport	1970	−0.039***	−0.040***	−0.039***	−0.064***	−0.070***	−0.074***
		(0.004)	(0.004)	(0.004)	(0.011)	(0.018)	(0.020)
Mother poorer health	1798	−0.027***	−0.025***	−0.026***	−0.024	−0.001	−0.000
		(0.004)	(0.005)	(0.005)	(0.017)	(0.031)	(0.035)
Household outcomes							
Partner's education	1456	0.651***	0.579***	0.573***	0.865***	0.819***	0.839***
(in years)		(0.022)	(0.028)	(0.029)	(0.081)	(0.128)	(0.145)
HH income *(log)*	2087	0.072***	0.068***	0.065***	0.108***	0.138***	0.140***
		(0.003)	(0.004)	(0.004)	(0.013)	(0.022)	(0.025)
Adolescents' schooling							
Daughters							
Academic track	952	0.072***	0.066***	0.065***	0.094***	0.099***	0.100***
		(0.005)	(0.006)	(0.006)	(0.009)	(0.017)	(0.019)
Sons							
Academic track	974	0.074***	0.067***	0.065***	0.091***	0.090***	0.089***
		(0.004)	(0.005)	(0.005)	(0.013)	(0.026)	(0.032)

Note: Average marginal effects of maternal education and robust standard errors (in parentheses) for Probit/OLS (1-3) and second stage of IV-Probit/two-stage least squares (2SLS) models. All regressions include controls for the child's and the mother's year of birth. Specifications 2 and 5 include additional fixed effects for the area the mother grew up and for the educational levels of the mother's parents, while specifications 3 and 6 control additionally for the grandparents' occupational prestige (ISEI score). * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

variable. The magnitude of the effects on the probabilities of the daughter being a smoker and of the daughter doing sports regularly change only marginally, but the effects become insignificant. This may be due to the substantial loss in precision.

The effects of maternal education on the daughter's health behavior may also work through a higher household income. Household income is measured by the logarithm of a five years average of household post-government income. The results in table 2.6 show that one more year of maternal education leads to an increase in household income of 14 per cent. To some extent this also accounts for assortative mating (Jepsen, 2005) because the mothers' partners are the principle earners in the majority of the households. When estimating a specification that includes the logarithm of household income as an additional control variable (table 2.5, specification 13), the estimated effects change only marginally.

Next, we investigate the child's schooling as potential channel. Usually after four years of primary school, the German school system selects children into one of three tracks: basic track (*Hauptschule*), intermediate track (*Realschule*), or academic track (*Gymnasium*). Pupils can only obtain the *Abitur* from academic track schools. The *Abitur* is the diploma usually required for enrolling into a German university. The IV estimates in table 2.6 suggest that one more year of maternal education increases the likelihood of the child pursuing the academic schooling track by about 10 percentage points. However, we do not find evidence for significant differences between sons and daughters.[26] Specification 14 (table 2.5) includes as additional control a binary variable indicating whether the adolescent attends an academic track school. As a result, the effect of maternal education on daughter's smoking behavior disappears while the effect on daughters doing sports regularly remains unchanged. Thus, maternal education seems to affect the daughter's smoking behavior by affecting schooling track. However, we cannot distinguish whether this is due to school quality (better understanding the risks of smoking, increased valuation of the future; see Fletcher and Frisvold, 2012) or due to peer group effects (lower share of smokers).

[26] Piopiunik (2011) instrumenting maternal education by changes in compulsory schooling legislation finds significant effects of maternal education on sons' but not on daughters' education. This paper analyzes the effect of maternal education at the lower tail of the education distribution, while we investigate effects over the whole distribution of maternal education.

2.6 Summary and discussion

Our paper investigates the effects of maternal education on child's health and health behavior in Germany. Using a rich survey panel data set (GSOEP), we analyze the effects on a wide range of outcomes for newborn and adolescent children. We estimate both Probit and IV-Probit models.

For newborns, we find a significant negative association between maternal education and the probability of preterm maturity. Although the effect on preterm maturity increases in size when estimating the IV-Probit model, the effect turns insignificant. It may be that the size of our newborns sample is not large enough to detect existing effects with the IV approach. We find no evidence for an effect of maternal education on low birth weight.

For adolescents, the IV approach suggests strong and significant effects on health-related behavior for daughters. One more year of maternal education is estimated to reduce the daughter's probability of smoking at age 18/19 by 7.4 percentage points and to increase the daughter's likelihood of doing sports at least once a week by 7.5 percentage points. However, we do not obtain significant effects of maternal education on sons' health behavior. For both sexes, we do not find any effects on child's physical health or overweight.

In line with previous research (e.g. Carneiro et al., 2013; Currie and Moretti, 2003), the significant estimates from the IV-Probit model exceed the corresponding estimates from the single equation Probit model. This may be attributed to three different reasons: measurement error in maternal education, unobserved variables leading to downward biased estimates in the Probit model, or the identification of local effects in the presence of effect heterogeneity when applying an IV-approach.

For our identification strategy, we do not rely on policy changes like previous studies. Instead, we present an IV approach that also works for cohorts unaffected by policy changes and for the limited sample sizes of common household panels. We argue that the mother's number of siblings is a valid instrument when conditioning on grandparental characteristics. Concerning the relevance of the instrument, we find all respective first stage F-statistics to exceed the critical value of 10. The estimation strategy seems not to suffer from a weak instruments problem. Regarding the validity of our instrument, we demonstrate the robustness of our IV estimates by sequentially introducing grandparental characteristics. The results are not substantially altered, when we include controls for grandparents' education, grandparent's occupational prestige and the size of the area where the mother grew up. Also the results do not change, when we control for further variables (like mother's health, health behavior, and fertility) that are possibly consequences of maternal education. Furthermore, the results are robust to only considering mothers with one, two or three siblings as well as to functional form assumptions of the first stage or distributional assumptions of the error terms.

Investigating possible channels of the estimated effects, our findings do not suggest that mother's health behavior, assortative mating, or household income explain the effects on daughter's health behavior. However, when including the child's schooling track as an additional control variable in our IV approach, the effect of maternal education on daughter's smoking behavior disappears. Thus, maternal education seems to affect child's health behavior by affecting schooling track. Even though early tracking is a particularity of the German schooling system, the mechanism at work (school quality or peer group effects) may also be relevant in other countries.

Public policy should take into account intergenerational links when thinking about optimal educational investments. There are persistent gains to be realized by increasing female education.

2.7 Appendix

Some variables on the mothers' and grandparents' level are affected by missing values; e.g. about 29 % of the children in the newborns sample have missing information on either the area where the mother grew up, grandparents' ISEI score, or grandfather's and grandmother's educational level. Omitting these cases will produce inefficient estimates, even if they are missing completely at random (MCAR; see Rubin, 1976). The estimates will be biased if the information is not MCAR but only missing at random (MAR). Under MAR the missingness depends on other observed variables, e.g. if mothers with fewer years of education know less about their parents.

Due to these effectiveness and unbiasedness considerations, we impute four variables relevant for our analysis: grandfather's and grandmother's educational level, grandparents' ISEI score and the area where the mother grew up. In case the information is missing, for all variables we first copy information provided by the mother's siblings. We impute missing values in the size of mother's area randomly conditional on the size of the mother's district of residence when she was interviewed in the GSOEP for the first time.

The other three variables are jointly imputed in four steps as follows. First, the educational levels of the grandparents are preliminarily imputed: If the level of education is missing for only one grandparent the information of the other grandparent is used. If the level of vocational training is available, the mode of level of education for each vocational training category is imputed. Second, we run a regression of the highest ISEI score of the grandparents (in most cases the grandfather's) score on sets of dummies for the grandfather's levels of vocational training and education, as well as dummies for the grandmother's levels of education and vocational training, dummies for the job position of the grandfather, controls for the birth decade of the grandfather and for each explanatory variable a dummy for missing values. These variables explain about 2/3 of the variance in grandparents' ISEI score. We exclude observations with missing information on all explanatory variables and do not impute any values for them.

Third, according to the regression results we predict values for those with missing information on the grandparents' ISEI score. We then add a random term drawn from the distribution of the regression residuals to maintain the variance of the dependent variable and to mimic the uncertainty of the imputation. Little and Rubin (2002, 60) refer to this procedure as stochastic regression imputation. Fourth, by means of multinomial logit models we regress the grandparents' educational level on the imputed grandparents' ISEI score, dummies for own vocational training levels and partner's education level. We use the predicted level of education for all those with missing information, including those with preliminarily imputed educational levels. In summary, we impute the grandfather's education for 7.5 % [7.0 %] of the adolescents [newborns], the grandmother's education for 6 % [5.5 %], the ISEI score for 44 % [16.5 %] and the size of the area the mother grew up for 1.5 % [9.8 %].

Table 2.7: Overview of previous IV-studies on the effect of maternal education on child's health

Author	Country	Dataset	Sample size	Instrument	Main findings
Currie and Moretti (2003)	USA	Vital Statistics Natality records	$> 670{,}000$	college openings	preterm birth: $\widehat{\beta}_1 = -0.01$ low birth weight (LBW): $\widehat{\beta}_1 = -0.01$
Carneiro et al. (2013)	USA	National Longitudinal Survey of Youth 1979	$> 3{,}600$ $> 8{,}000$	variation in mother's schooling costs	no effects on overweight (age 7-8, 12-14) no effects on LBW
Chevalier and O'Sullivan (2007)	UK	National Child Development Study	$> 8{,}000$	compulsory schooling reform	no effect on LBW birth weight: $\widehat{\beta}_1 = 0.74g$
Lindeboom et al. (2009)	UK	National Child Development Study	$> 11{,}000$	compulsory schooling reform	no effects on birth weight, LBW and overweight (at ages 7, 11 and 16)
McCrary and Royer (2011)	USA	natality data for TX and CA	$> 320{,}000$	age-at-school-entry policies	no effects on LBW and preterm birth

Note: The table presents an overview of previous IV-studies on the effect of maternal education on child's health in developed countries. Specifically, the table provides information about country, data set, sample size, applied instrument and main findings of the studies that analyze similar outcomes. $\widehat{\beta}_1$ refers to the estimated effect of one more year of maternal education, see equation (2.3).

Chapter 3

Bayesian procedures as a numerical tool for the estimation of dynamic discrete choice models†

3.1 Introduction

Accounting for unobserved heterogeneity is important when estimating non-linear models. Numerous studies document that discrete choice models without unobserved heterogeneity require either very strong and often implausible assumptions or lead to biased estimates of central parameters. In particular, this is true for dynamic models which analyze the role of state dependence in the behavior of agents because in these models it is necessary to disentangle true state dependence from individual specific effects, see for example Heckman (1981a, 1981b).[27] For a recent study in labor economics see e.g. Prowse (2010), for an example from marketing see Dube et al. (2010). The structure of unobserved heterogeneity in discrete choice models can be complex and therefore it is often necessary to allow for a general specification with potential correlations of the different processes. Estimation of such models often is computationally expensive and numerical problems with classical estimation procedures may arise (e.g. non-convergence or convergence at a local maximum of the opti-

† This chapter is based on joint work with Peter Haan and Arne Uhlendorff, see Haan et al. (2012).

[27] The same applies for hazard rate models which try to separate duration dependence from dynamic selection due to unobserved characteristics, see e.g. Lancaster (1990) and van den Berg (2001) for overviews.

mization algorithm). These problems are particularly severe when unobserved heterogeneity is assumed to follow a less convenient distribution - such as a log-normal distribution - rather than a multivariate normal distribution which is the standard assumption. Train (2009) points out that in the classical approach locating the maximum of the likelihood is considerably more difficult with log-normal distributions. And, even if a maximum is found, it may happen that the Hessian is singular at this point.

However, economic theory clearly indicates that in many applications the assumption of a normal distribution of unobserved effects is not appropriate since this potentially leads to counterintuitive effects, e.g. negative preferences for income of consumption or positive price effects of normal goods. Bayesian procedures, such as Markov Chain Monte Carlo (MCMC) methods serve as an important alternative for the estimation of non-linear models with unobserved heterogeneity. Since the Bayesian MCMC estimator does not involve maximization of a likelihood function, the numerical problems of classical procedures such as maximum simulated likelihood (MSL) do not arise.[28] Consistency and efficiency can be achieved under more relaxed conditions.[29]

The use of MCMC methods in the analysis of dynamic discrete choice models has become more popular in recent years. Fitzenberger et al. (2010) and Troske and Voicu (2010) are two examples in the context of dynamic labor supply models, whereby Fitzenberger et al. (2010) focus on the impact of training incidence and duration on employment outcomes and Troske and Voicu (2010) analyze interdependencies between fertility choices and labor supply behavior of females. Imai et al. (2009) propose a method applying MCMC methods combined with dynamic programming for structural estimation of dynamic discrete choice models. However, the majority of applied micro-econometric studies estimating discrete choice panel data models with lagged states applies classical maximum likelihood approaches, see e.g. Akay (2009), Haan (2010) and Prowse (2010).

[28] In the context of static mixed logit models Regier et al. (2009) demonstrate that MSL estimation converge at a local maximum when unobserved effects follow a log-normal distribution. Therefore - similar to our strategy for dynamic models in this paper - they suggest to use Bayesian methods for the estimation of the static model.

[29] Zellner and Rossi (1984) were the first to apply Bayesian procedures for a logit model using importance sampling, and starting with Zeger and Karim (1991) various MCMC methods have been developped for the estimation of discrete choice models, see Albert and Chib (1993) and McCulloch and Rossi (1994) for probit models, and Allenby and Lenk (1994) and Allenby (1997) for logit models.

The contribution of our paper is to compare the numerical performance of Bayesian estimation procedures with MSL estimation of dynamic discrete choice models and to provide evidence about the advantages of Bayesian methods for practitioners. We consider an application of a dynamic discrete choice model of female labor supply with three distinct states and estimate dynamic mixed logit models. Our analysis is based on longitudinal data from the German Socioeconomic Panel (GSOEP).[30]

The empirical comparison leads to two important conclusions which are highly relevant for practitioners estimating dynamic discrete choice models with unobserved heterogeneity. First, when considering a multivariate normal distribution for the unobserved heterogeneity both approaches, the MCMC estimator and the MSL estimation, yield almost identical results. This shows that for a finite sample of the size which is typical for common household panels, our findings are in line with the asymptotic results of the Bernstein-von Mises Theorem. Hence, the Bayesian estimates can be given a classical interpretation. Second, we show that when imposing distributional assumptions about the unobserved heterogeneity which are consistent with economic theory, e.g. log-normally distributed consumption preferences, only the Bayesian method provides reliable estimates. The classical estimator does not converge because it fails to find an increase of the likelihood function after a few iterations. This is the case even for starting values that have been set close to the maximum of the likelihood function. With log-normally distributed consumption preferences the respective coefficient decreases significantly by about 20%. In our example, this translates not into significantly different labor supply elasticities for the population mean.

The paper is organized as follows. After a general comparison between Bayesian and classical estimation of discrete choice models with unobserved heterogeneity, we outline our MCMC procedure. Then, we describe the dynamic labor supply model, the detailed specification and the data for our specific application. We compare estimation results of the dynamic labor supply model with different specifications of unobserved heterogeneity using i) classical and ii) Bayesian estimation procedures. A final section summarizes our findings.

[30]From a practitioner's perspective, it seems more interesting to assess the numerical performance of the Bayesian and maximum likelihood procedures with real world data than with a simulated data set because the distributional assumptions of a simulated data set often appear a bit artificial.

3.2 Classical versus Bayesian estimation of discrete choice models

In the following, we describe the classical and the Bayesian estimation procedure of discrete choice models. For convenience we use a similar notation as in Train (2009). We start with the classical perspective.

3.2.1 Maximum Simulated Likelihood Estimation

Let's assume that each individual i maximizes period t's utility U_{ijt} by making choice $j \in 1, ..., J$. The probability that individual i makes a series of choices $\mathbf{j} = (j_1, ..., j_T)$ in the T observed time periods is $P_{i\mathbf{j}} = Prob(U_{ijt} > U_{ikt}, \forall j \neq k, \forall j \in \mathbf{j}, \forall t = 1, ..., T)$. Allowing for random taste variation, we can set up a random utility model $U_{ijt} = \beta_i x_{ijt} + \epsilon_{ijt}$ where x_{ijt} is a vector of covariates, β_i is an individual-specific (random) coefficient vector, and ϵ_{ijt} is an idiosyncratic error term. We denote the conditional probability of i making a series of choices $\mathbf{j}$ as $L_{i\mathbf{j}}(\beta_i)$. Then, we can write the unconditional probability $P_{i\mathbf{j}} = \int L_{i\mathbf{j}}(\beta_i) f(\beta_i|\theta) d\beta_i$ where $f(\beta_i|\theta)$ is the density function describing the distribution of β_i conditional on the parameters θ of the distribution. Thus, we integrate over all possible states of β_i. Usually, the integral must be approximated by simulation methods. Assuming that ϵ_{ijt} follows an extreme value distribution, the model becomes a mixed logit model. McFadden and Train (2000) show that any random utility model can be approximated to any degree of accuracy by a mixed logit model if the density function $f(\beta_i|\theta)$ is chosen appropriately.

The probability $P_{i\mathbf{j}}$ can be approximated by simulation procedures like the method of simulated likelihood.[31] The idea is that taking draws β_i^r from $f(\beta_i|\theta)$ allows calculating $L_{i\mathbf{j}}(\beta_i^r)$. The simulated probability is then $\hat{P}_{i\mathbf{j}} = \frac{1}{R} \sum_{r=1}^{R} L_{i\mathbf{j}}(\beta_i^r)$ where R is the number of draws. The simulated log likelihood function becomes

$$SLL = \sum_{i=1}^{N} \sum_{\mathbf{j} \in \mathbf{J}} d_{i\mathbf{j}} log \hat{P}_{i\mathbf{j}} \tag{3.1}$$

$d_{i\mathbf{j}}$ equals 1 if individual i makes the series of choices $\mathbf{j}$ and zero otherwise. $\mathbf{J}$ is the set of all possible series of choices. The maximum simulated likelihood estimator (MSLE) $\hat{\theta}^{MSL}$ of θ is the value that maximizes the SLL. The max-

[31] See Train (2009) for more details on this method as well as alternative approaches like the method of simulated moments and the method of simulated scores.

imum of the SLL can be found using numerical methods. However, problems may arise if the optimization algorithm does not converge or converges at a local maximum of the likelihood function. The MSLE is consistent if R rises at any rate with $\sqrt{N}$. It is also efficient if R rises faster than $\sqrt{N}$. Note that for R fixed the MSLE is inconsistent. For the asymptotic distribution of $\hat{\theta}^{MSL}$, we have $\sqrt{N}(\hat{\theta}^{MSL} - \theta) \xrightarrow{d} N(0, -H^{-1})$. -H is the information matrix. Thus, the MSLE is consistent, efficient and asymptotically normal if R rises faster than $\sqrt{N}$.

3.2.2 Bayesian Estimation

When estimating a discrete choice model with unobserved heterogeneity, there is a Bayesian analog to the classical procedures such as the method of MSL. From a Bayesian perspective the prior beliefs of the researcher about the parameters θ are represented by a prior distribution $k(\theta)$. Given the observed series of choices Y of the N decision makers, there is a density $K(\theta|Y)$ that summarizes the information about θ that is provided by Y. This density is called posterior distribution. The relationship between $k(\theta)$ and $K(\theta|Y)$ is given by Bayes rule. If $P_{i\mathbf{j}}(\theta)$ is the probability that decision maker i makes the series of choices $\mathbf{j}$, the probability of observing the sample is given by $L(Y|\theta) = \prod_{i=1}^{N} P_{i\mathbf{j}}(\theta)$. By Bayes rule $K(\theta|Y)L(Y) = L(Y|\theta)k(\theta)$ where $L(Y) = \int L(Y|\theta)k(\theta)d\theta$. It follows that $K(\theta|Y) = \frac{L(Y|\theta)k(\theta)}{L(Y)}$.

The relationship between the classical and the Bayesian perspective is stated by the Bernstein-von Mises Theorem. A formal outline of the theorem can be found in Train (2009). First, the posterior distribution of θ converges to a normal distribution with variance $\frac{-H^{-1}}{N}$ as the sample size increases. -H is the information matrix being well-known from classical statistics. Thus, the parameters' posterior distribution converges to the sampling distribution of the maximum likelihood estimator. Second, the mean of the posterior distribution $\bar{\theta} = \int \theta K(\theta|Y)d\theta$ converges to the maximum of the likelihood function.[32] As a consequence, $\bar{\theta}$ is asymptotically equivalent to the MSLE. Since the Bayesian MCMC-estimator does not involve maximization of a likelihood function, the numerical problems of classical procedures do not arise. Standard errors can easily be calculated recognizing that the variance of the posterior distribution of θ is the asymptotic sampling variance of the MSLE. Note that

[32] This follows directly from the symmetry of the normal distribution. As has been pointed out, the posterior distribution of the parameters is asymptotically normal.

this is a classical perspective on Bayesian estimates. Simulation methods allow an approximation of $\bar{\theta}$ by taking R draws from the posterior distribution. Then, $\hat{\theta}^{MCMC} = \frac{1}{R}\sum_{r=1}^{R}\theta^r$ is the simulated mean where θ^r is the rth draw from $K(\theta|Y)$. Since the standard errors of the estimates correspond to the standard deviation of $K(\theta|Y)$, they can be simulated by taking the standard deviation of the R draws. For a fixed number of draws, the simulated mean of the posterior distribution is consistent and asymptotically normal. We have $\sqrt{N}(\hat{\theta}^{MCMC} - \theta) \xrightarrow{d} N(0, -H^{-1})$ as for the MSLE. If the number of draws increases at any rate with sample size, the estimator is efficient. Thus, consistency and efficiency can be achieved under more relaxed conditions than for the MSLE. Draws from the posterior distribution $K(\theta|Y)$ can be taken by MCMC methods.

3.2.3 MCMC procedure

We follow an estimation procedure that has been developed by Allenby and Lenk (1994) and Allenby (1997), and generalized by Train (2001). The approach relies on diffuse priors and applies Gibbs sampling together with the more general, but computationally more expensive, Metropolis-Hastings algorithm (Metropolis et al. (1953) , Hastings (1970)). Gibbs sampling allows taking draws from the posterior distribution of subsets of the parameters conditional on the other parameters. We specify diffuse priors for the mean b and variance-covariance matrix W of the individual-level coefficients β_i: k(b) is $N(b_0, S_0)$ with S_0 being extremely large (i.e. flat prior) and k(W) is inverted Wishart IW(F,I) where F is the number of random coefficients and I is an F-dimensional identity matrix. In empirical applications, we often have no clear idea about the prior distribution. Therefore, it seems to be reasonable to assume an uninformative prior. If the utility that individual i obtains in period t from choosing alternative j is $U_{ijt} = \beta_i x_{ijt} + \epsilon_{ijt}$, drawing from the posterior distribution $K(b, W|Y)$ is fast and convenient when β_i is considered as parameters along with b and W (see Train (2009) for more details). Then, the posterior distribution is

$$K(b, W, \beta_i|Y) \propto L_{\mathbf{ij}}(\beta_i)\phi(\beta_i|b, W)k(b, W), \forall i \tag{3.2}$$

where ϕ is the normal density function with mean b and variance-covariance matrix W. Since some of the coefficients are kept fixed, this requires an additional layer of the Gibbs sampling. The fixed coefficients are denoted α and the utility becomes $U_{ijt} = \alpha z_{ijt} + \beta_i x_{ijt} + \epsilon_{ijt}$. Then, the individual-level coefficients β_i must be drawn conditional on α and vice versa. The prior distribution of α, k(α), is also assumed to be normal and flat. Then, we get the following four posterior distributions for the four sets of parameters β_i, b, W, and α:

1. $K(\beta_i|\alpha, b, W) \propto L_{\mathbf{ij}}(\alpha, \beta_i)\phi(\beta_i|b, W)$, $\forall i$. The Metropolis-Hastings algorithm is used to draw from this posterior distribution.

2. $K(b|W, \beta_i)$ is $N(\frac{1}{N}\sum_i \beta_i, \frac{W}{N})$, $\forall i$.

3. $K(W|b, \beta_i)$ is $IW(K+N, \frac{KI+N\bar{S}}{K+N})$, $\forall i$, where F is the number of random coefficients, I is an F-dimensional identity matrix and $\bar{S} = \frac{\sum_i(\beta_i - b)(\beta_i - b)'}{N}$.

4. $K(\alpha|\beta_i) \propto \prod_i L_{\mathbf{ij}}(\alpha, \beta_i)$, $\forall i$. The Metropolis-Hastings algorithm is used to draw from this posterior distribution.

Setting the initial values for all parameters to 0.1 and using Gibbs sampling to draw from the posterior distribution of the four sets of parameters, it takes a number of iterations until the draws converge to draws from the posterior distribution. The draws prior to convergence must be discarded. In our empirical application we discard the first 20,000 draws for burn-in and use the following 10,000 draws for the actual estimation. We always check on convergence by increasing the number of draws used for burn-in and, then, comparing the results. If results do not change even with a substantial increase of the draws used for burn-in, this suggests that convergence has been achieved. The acceptance rate in the Metropolis-Hastings algorithm is set to be 0.3.[33] Since the draws are serially correlated, we follow the suggestion of Train (2009) and save only every tenth draw reducing the correlation of the retained draws by an order of ten.

3.3 Economic model and specification

We specify a dynamic discrete choice model with random coefficients to model labor supply of married females. Our approach assumes that current labor supply is causally affected only by last period's labors supply, thus following a first-order Markov process, see, e.g. Keane and Wolpin (2001) or Haan (2010). We differentiate between three distinct states, $j \in 1, 2, 3$: full-time, part-time

[33] Gelman et al. (1995) find that the optimal acceptance rate is about 0.44 if x_{ijt} contains only one variable and decreases with the number of variables in x_{ijt} towards 0.23.

and non-employment. The median hours of work per week are 13.5 for part-time employees and 38 for full-time employees. We focus on married females in prime working age, defined as 25-59 years. The wife maximizes her utility conditional on her husband's behavior which is taken as given. Collective bargaining as suggested by e.g. Chiappori (1988) is not considered. Each period t, wife i receives a utility flow U_{ijt} depending on her choice of labor supply category j. U_{ijt} is a function of period t's leisure time and household consumption as well as of the previous period's labor market state. Since our model does not account for saving behavior, household consumption equals net household income. Net household income is simulated using the micro simulation model STSM (Steiner et al. (2008)), for more details, see below. We ensure concavity of the preferences by assuming a Cobb-Douglas functional form of the utility function:

$$U_{ijt} = \beta_{it}^{L} Log L_{jt} + \beta_{i}^{I} Log I_{ijt} + \epsilon_{ijt} \tag{3.3}$$

$$\beta_{it}^{L} = b_{i}^{L} + b^{C} C_{it} + b^{E} E_{it} + b_{10}^{Z} Z_{1i0} + b_{20}^{Z} Z_{2i0} + b_{1i}^{Z} Z_{1it-1} + b_{2i}^{Z} Z_{2it-1} \tag{3.4}$$

where L_{jt} is wife's leisure time and I_{ijt} is household consumption (i.e. net household income). Z_{1it-1} and Z_{2it-1} are dummy variables indicating the state which has been chosen in the previous period; Z_{1it-1}: non-employment, Z_{2it-1}: part-time. Analogously Z_{1i0} and Z_{2i0} are variables indicating the employment state of the first observed period (2001). Thus, we follow Wooldridge (2005) by modelling the distribution of unobserved heterogeneity conditional on the first observed state to account for the nonrandomness of the initial state.[34] The lagged dependent variables and the initial state are modeled as taste shifters of the leisure time in the current period, L_{jt}. Similarly, other individual specific characteristics are included as taste shifters for leisure. C_{it} is a dummy variable for children aged 0-3 years in the household and E_{it} is a dummy variable indicating residence in East Germany. ϵ_{ijt} is an i.i.d. error term that follows an extreme value distribution. Note that in the above specification β_i^I, b_i^L, b_{1i}^Z, and b_{2i}^Z are individual specific coefficients which introduce unobserved preference heterogeneity to the model. Distributional assumptions for these random coefficients must be imposed, and a standard approach in the literature is to assume a multivariate normal distribution for the unobserved components. This leads to a mixed logit model (McFadden and Train (2000)).

[34]See Akay (2011) for the performance of the method suggested by Wooldridge (2005).

We estimate various specifications of the model. In a first specification, we neglect unobserved preference heterogeneity and assume that the variance-covariance matrix of the individual specific coefficients is zero. This results in a simple conditional logit model. The second specification only assumes two random coefficients, for b_i^L and β_i^I that are uncorrelated. Specification 3 extends specification 2 allowing the two random coefficients to be correlated. In specification 4, we estimate the full model allowing for four random coefficients (also on the state dependence) and for a free correlation structure. Specifications 2 to 4 assume a multivariate normal distribution of the random coefficients which is a standard approach in the literature. While this makes sense for the leisure preferences, it is problematic for consumption preferences, see e.g. van Soest (1995). Assuming a normal distribution of unobserved heterogeneity in consumption preferences predicts negative preferences for some individuals. In our example, about 15 % of the individuals are predicted to have negative consumption preferences when estimating specification 4. This is not plausible from an economic point of view. Therefore, we estimate a fifth specification that restricts consumption preferences to be positive by assuming them to be log-normally distributed. This can be easily achieved by setting β_i^I equal to $e^{b_i^I}$ and estimating b_i^I under the assumption that b_i^I is normally distributed. Thus, the utility function becomes:

$$U_{ijt} = \beta_{it}^L LogL_{jt} + e^{b_i^I} LogI_{ijt} + \epsilon_{ijt} \tag{3.5}$$

If b_i^I is normally distributed, $e^{b_i^I}$ follows a log-normal distribution. We can obtain draws from the posterior distribution of the log-normally distributed β_i^I by transforming the draws from the posterior distribution of b_i^I with $e^{(\cdot)}$. The variance of the log-normally distributed coefficient can be estimated transforming draws from the posterior distribution of the mean and the variance of b_i^I to draws from the posterior distribution of the variance of β_i^I. Note that $var(\beta_i^I) = e^{(mean(b_i^I)+var(b_i^I)/2)} * (e^{var(b_i^I)} - 1)$. Specification 5 also allows for a free correlation structure.

The first specification is estimated by the method of maximum likelihood using an analytic gradient in the optimization algorithm. Specifications 2 to 4 are estimated by the method of MSL using again an analytic gradient.[35] For

[35]The simulation of the choice probabilities is based on 200 Halton draws. Estimating the model with other choices of R has shown that 200 Halton draws seems to be a lower bound to the number of draws required for the MSLE to have good statistical properties in our finite sample.

comparison, we estimate specification 4 as well using the Bayesian MCMC estimator.[36] Specification 5 could not be estimated by the MSL-estimator because the optimization algorithm does not converge because it fails to find an increase of the likelihood function after a few iterations. This is the case even for starting values that have been set close to the maximum of the likelihood function. Train (2009) points out that in the classical approach locating the maximum of the likelihood is considerably more difficult with log-normal distributions. And, even if a maximum is found, it may happen that the Hessian is singular at this point. However, specification 5 can be estimated easily by the Bayesian procedures. Since the Bayesian MCMC estimator does not involve maximization of a likelihood function, the numerical problems of the classical approach do not arise.

3.4 Data

The empirical analysis is based on data from the German Socio-Economic Panel (GSOEP). The GSOEP started in 1984 and annually collects information at the household and individual levels (Wagner et al. (2007)). We construct a balanced panel of married females covering the years 2001 to 2008. The sample is restricted to married females in prime working age defined as 26-59 years whose labor supply may be considered as flexible, thus excluding pensioners, self-employed, and people in institutions. The final sample consists of 1,598 wives. The GSOEP contains detailed information about employment behavior, as well as other individual and household characteristics. As mentioned above, we differentiate between non-employment, part-time and full-time work. The median hours of work per week are 13.5 for part-time employees and 38 for full-time employees. Leisure time is defined as 80 minus median hours of work per week of the employment category.

Table 3.1 contains some descriptive statistics including the distribution of the working alternatives. Net household income is simulated for each employment category using the micro simulation model STSM (Steiner et al. (2008)). Net household income is a complex nonlinear function of the gross wages of

[36] We only compare the estimation of specification 4 with both estimation procedures, the comparison of the less flexible specifications 2 and 3 lead to the same results and are therefore omitted.

Table 3.1: Descriptive statistics

Variable	Mean	Std.	Min	Max
Age	44.6	7.1	26	59
Children from 0-3 years	0.054	0.227	0	1
German nationality	0.932	0.251	0	1
East Germany	0.232	0.422	0	1
Non-employment	0.21	0.407	0	1
Part-time	0.263	0.44	0	1
Full-time	0.527	0.5	0	1

both partners.[37] The STSM allows taking into account the complexity of the German tax and transfer system. Simulation is based on detailed information on household characteristics as contained in the GSOEP. Other explanatory variables entering directly our labor supply model are a regional dummy indicating whether a household is living in eastern or western Germany and a binary variable providing information on young children aged three or less in the household.

3.5 Results and discussion

3.5.1 Parameter estimates

Table 3.2 contains the parameter estimates for the five specifications that have been described above. Assuming no preference heterogeneity (see specification 1) seems to result in a substantial underestimation of the income coefficient, β_i^I. When allowing for correlation between random coefficients, the size of β_i^I increases significantly in comparison to the specification without correlated random coefficients (compare specifications 2 und 3). This finding supports previous studies, e.g. Prowse (2010) that specifications without unobserved heterogeneity lead to biased estimates of central structural parameters. Because of the multiple interactions of the leisure term the interpretation of coefficient b_i^L is not straight forward. However we find again a substantial difference in this estimated coefficient between models with and without unobserved heterogeneity. Neglecting heterogeneity in the coefficients of state dependence, b_{1i}^Z and b_{2i}^Z, results in significantly smaller estimates of these coefficients (compare specifications 3 and 4). This misspecification is in particular relevant for the

[37]Note, for women not employed in the month preceding the interview, gross hourly wages are estimated by applying a two-stage estimation procedure with a Heckman sample selection correction.

interpretation of state dependence (Heckman (1981c)) and the prediction of short term labor supply elasticities. Looking at the information criteria AIC and BIC, specification 4 appears to have the better trade-off between flexibility and efficiency in comparison to specifications 1-3. Both a free correlation structure of the random coefficients and accounting for heterogeneity in the coefficients of state dependence improve the fit of the model. It appears that neglecting the correlation between random coefficients leads to an underestimation of the variances. The estimated covariances indicate a positive correlation between income and leisure preferences. To summarize we find in line with the previous literature that it is important to account for unobserved heterogeneity in a flexible and general way.

However, a more detailed comparison of the results leads to important conclusions for practitioners related to the Bayesian estimations procedures that go beyond the standard findings. First, when considering a multivariate normal distribution for the unobserved heterogeneity (e.g. specification 4) both approaches, the MCMC estimator and the MSL estimation, yield almost identical results. The estimated means of the posterior distribution of the parameters and of the variance-covariance matrix of the random coefficients differ only marginally from the parameter estimates of the MSL estimation. This shows that for a finite sample of the size which is typical for common household panels, the Bernstein-von Mises Theorem holds and the Bayesian estimates can be given a classical interpretation which is central for the interpretation of the results.

Second, we show that when imposing distributional assumptions about the unobserved heterogeneity which are more in line with economic theory only the Bayesian method provides reliable estimates. In particular, restricting consumption preferences to be positive by assuming them to be log-normally distributed yields an estimate of the respective coefficient, β_i^I, that is significantly smaller than in the case of normally distributed consumption preferences (compare specifications 4 and 5). The distribution of a normally distributed β_i^I is shifted to the right when fitting the model such that a smaller share of individuals is predicted to have negative consumption preferences. Restricting β_i^I to be positive is necessary to avoid an overestimation of the coefficient. The classical estimator does not converge because it fails to find an increase of the likelihood function after a few iterations. This finding even holds for starting values that have been set close to the maximum of the likelihood function. With log-normally distributed consumption preferences the respective coefficient de-

Table 3.2: Parameter estimates

Parameters	ML (spec. 1)	MSL (spec. 2)		MSL (spec. 3)		MSL (spec. 4)		MCMC (spec. 4)		MCMC (spec. 5)	
	value	value	var.	value	var.	value	var.	mean	var.	mean	var.
b^Z_{10}	3.30	9.54		10.75		5.75		5.88		4.03	
	(0.19)	(0.58)		(0.59)		(0.66)		(0.55)		(0.32)	
b^Z_{20}	2.05	5.92		6.25		4.02		4.05		2.81	
	(0.15)	(0.42)		(0.43)		(0.45)		(0.38)		(0.20)	
b^C	2.75	4.73		5.24		5.06		4.90		4.17	
	(0.29)	(0.41)		(0.44)		(0.45)		(0.42)		(0.35)	
b^E	-1.36	-2.58		-2.53		-2.10		-2.20		-1.71	
	(0.17)	(0.35)		(0.39)		(0.31)		(0.31)		(0.26)	
b^Z_{1i}	9.34	5.64		5.60		11.07	51.34	11.02	55.87	12.27	67.72
	(0.23)	(0.32)		(0.32)		(0.72)	(7.80)[1]	(0.58)	(8.25)	(0.53)	(8.55)
b^Z_{2i}	4.54	2.46		2.32		4.18	5.87	4.07	5.14	4.87	6.91
	(0.15)	(0.21)		(0.21)		(0.37)	(1.78)[1]	(0.30)	(1.21)	(0.23)	(1.04)
b^L_i	-5.17	-4.31	8.51	-3.39	51.93	-3.30	18.21	-3.22	19.61	-4.21	7.63
	(0.15)	(0.29)	(1.40)[1]	(0.37)	(6.14)[1]	(0.33)	(3.74)[1]	(0.30)	(3.83)	(0.24)	(1.81)
β^I_i	1.89	7.54	26.61	10.38	167.95	9.39	92.46	9.41	89.49	6.97	81.38[2]
	(0.19)	(0.53)	(4.12)[1]	(0.69)	(22.94)[1]	(0.78)	(15.18)[1]	(0.64)	(14.96)	(0.54)	(24.52)
random coeff.	no	2		2		4		4		4	
free correlation	no	no		yes		yes		yes		yes	
log-normal β^I_i	no	no		no		no		no		yes	
ML / MSL	-6393.1	-6245.8		-6137.5		-6052.5		-		-	
AIC	12802.3	12511.5		12297.0		12140.9		-		-	
BIC	12845.3	12565.3		12356.1		12237.7		-		-	

Standard errors / standard deviations of the posterior distributions are given in parentheses.

[1] For numerical reasons, we estimate the elements of the Cholesky matrix of the variance-covariance matrix with the method of MSL. The standard errors of the derived variances, then, are approximated on the basis of 200 draws taken from the asymptotically normal sampling distribution of the elements of the Cholesky matrix.

[2] The variance of the log-normally distributed coefficient has been estimated transforming draws from the posterior distribution of the mean and the variance of b^I_i to draws from the posterior distribution of the variance of β^I_i. Note that $\beta^I_i = e^{b^I_i}$.

creases significantly by about 20 % and the derived labor supply elasticities become slightly smaller.

3.5.2 Labor supply elasticities

For an economic interpretation of the female labor supply behavior it is necessary to compute labor supply elasticities. Therefore, we use the estimates for different specifications to derive elasticities for a permanent shock to gross wage. We only derive elasticities for the population mean. In particular we calculate the relative change in the labor supply behavior which result from a permanent relative increase in female gross wages. We predict for 10 consecutive periods of time for both a baseline scenario without a shock and a scenario where a permanent wage shock occurs in the first period, i.e. wages are 10% higher in the second scenario. Then, we use the differences between the predictions of the two scenarios to derive the labor supply elasticities. All predictions are based on the posterior distribution of the model parameters.[38] This allows investigating the dynamics of female labor supply. For a detailed description of the calculation see Haan (2010) and Haan and Uhlendorff (2012).

Table 3.3 contains the labor supply elasticities. The elasticities for period 1 can be interpreted as short term elasticities, while the elasticities for period 10 can be interpreted as long term elasticities. Confidence intervals are estimated at the 0.95 significance level. The confidence intervals for the elasticities derived from the MCMC estimation have been computed directly using the draws from the posterior distribution of the model parameters. We approximate confidence intervals for the elasticities derived from the ML and MSL estimation by taking 200 draws from the sampling distribution of the model parameters and deriving the elasticities for each of these draws. Unlike the Bayesian confidence intervals, this approach must assume that the asymptotically normal sampling distribution of the parameters holds for our finite sample. Note that the interpretation of Bayesian confidence intervals differs from the interpretation of classical confidence intervals. The Bayesian confidence interval represents the belief that the true elasticity lies within the interval with a probability of 0.95.

The estimated elasticities differ depending on whether random coefficients are allowed for. It results from the underestimation of the income coefficient in the conditional logit model that neglecting preference heterogeneity seems to

[38]This is natural for predictions based on Bayesian estimation. For predictions based on MSL estimation, we have to compute the choice probabilities conditional on each individual's series of choices by a procedure that involves simulation within simulation.

Table 3.3: Labor supply elasticities

Period	ML (spec. 1)	MSL (spec. 2)	MSL (spec. 3)	MSL (spec. 4)	MCMC (spec. 4)	MCMC (spec. 5)
1	0.054	0.143	0.165	0.141	0.143	0.124
	[0.044,0.064]	[0.127,0.158]	[0.155,0.176]	[0.124,0.157]	[0.127,0.159]	[0.109,0.141]
2	0.080	0.184	0.207	0.190	0.193	0.163
	[0.065,0.096]	[0.165,0.203]	[0.195,0.219]	[0.170,0.211]	[0.174,0.214]	[0.144,0.186]
3	0.096	0.199	0.221	0.213	0.219	0.187
	[0.077,0.114]	[0.179,0.219]	[0.208,0.233]	[0.192,0.234]	[0.196,0.243]	[0.164,0.213]
4	0.105	0.205	0.225	0.225	0.234	0.202
	[0.085,0.126]	[0.184,0.225]	[0.212,0.239]	[0.204,0.247]	[0.210,0.261]	[0.177,0.230]
5	0.110	0.207	0.227	0.232	0.245	0.212
	[0.089,0.131]	[0.186,0.228]	[0.213,0.241]	[0.211,0.254]	[0.219,0.273]	[0.186,0.242]
6	0.113	0.208	0.228	0.236	0.252	0.220
	[0.092,0.135]	[0.187,0.229]	[0.214,0.241]	[0.215,0.258]	[0.225,0.281]	[0.193,0.251]
7	0.115	0.208	0.228	0.239	0.258	0.226
	[0.094,0.137]	[0.187,0.229]	[0.214,0.242]	[0.217,0.260]	[0.230,0.288]	[0.198,0.257]
8	0.117	0.208	0.228	0.240	0.262	0.231
	[0.095,0.138]	[0.187,0.229]	[0.214,0.242]	[0.219,0.262]	[0.234,0.293]	[0.201,0.263]
9	0.117	0.208	0.228	0.241	0.265	0.235
	[0.095,0.138]	[0.187,0.229]	[0.214,0.242]	[0.219,0.263]	[0.237,0.297]	[0.205,0.268]
10	0.118	0.208	0.228	0.242	0.268	0.239
	[0.096,0.139]	[0.187,0.229]	[0.214,0.242]	[0.220,0.263]	[0.239,0.300]	[0.207,0.272]

We calculate confidence intervals at the 5 percent significance level. Note that the interpretation of the Bayesian confidence intervals differs from the interpretation of the classical confidence intervals.

induce a substantial downward bias to the estimated elasticities. For example, the conditional logit model predicts a long term labor supply elasticity of about 0.12 as opposed to 0.24 for specification 4 (MSL). The difference is significant. The predicted elasticities do only differ marginally depending on whether the model has been estimated by the method of MSL or by the Bayesian MCMC estimator. This reflects the very similar parameter estimates of both estimation approaches. The smaller income coefficient of specification 5 translates not into significantly different labor supply elasticities for the population mean.

3.6 Conclusion

This analysis provides evidence for the advantage of using Bayesian estimation procedures instead of classical maximum likelihood estimation for the estimation of dynamic discrete choice models. These models usually require a general specification of unobserved preference heterogeneity and therefore often relatively complex estimation routines need to be applied. We estimate a dynamic discrete choice model of female labor supply with three distinct states and different specifications of unobserved heterogeneity.

The empirical comparison leads to two important conclusions which are highly relevant for practitioners estimating dynamic discrete choice models with unobserved heterogeneity. First, when considering a multivariate normal distribution for the unobserved heterogeneity both approaches, the MCMC estimator and the MSL estimation, yield almost identical results. This shows that for a finite sample of the size which is typical for common household panels, our findings are in line with the asymptotic results of the Bernstein-von Mises Theorem. Hence, the Bayesian estimates can be given a classical interpretation. The second finding demonstrates the advantage of using Bayesian estimation procedures. We show that when imposing distributional assumptions which are consistent with economic theory, e.g. log-normally distributed consumption preferences, the Bayesian method performs well and provides reasonable estimates, while the MSL estimator does not converge. These results indicate that Bayesian procedures can be a beneficial tool for the estimation of dynamic discrete choice models.

Chapter 4

A life-cycle perspective on the health-related risks of consumption and old age poverty

4.1 Introduction

The strong association between health and socio-economic status is a very robust finding and is discussed by a large body of literature (see David and Lleras-Muney (2012) or Grossman (2006) for overviews). While there is dissent about whether or not and to what extent income affects health in developed countries, there is strong evidence that a substantial share of income inequality can be explained by health (Deaton (2003)). In particular, an individual's health status is found to be one of the main determinants of both labor market participation and early retirement. Furtheremore, unemployment and early retirement reduce an individual's expected life-cycle income with potentially long-lasting impacts on wealth accumulation. As has been pointed out by Deaton (2003), this is due to the fact that individuals cannot fully insure their earnings against health risks. Since health shocks affect the dynamics of labor market participation, early retirement, and wealth accumulation, a structural econometric analysis of health and economic outcomes should take into account the relevant processes and the individuals' expectations about the consequences of their decisions. This paper proposes a rich life-cycle model in order to investigate health-related consumption and poverty risks.

While the paper follows a tradition of dynamic retirement models (e.g. Rust and Phelan (1997); French (2005); van der Klaauw and Wolpin (2008); French and Jones (2011); and Haan and Prowse (2011)), it adds to the literature in several ways. First, the study demonstrates the good performance of an extension of the Expectation-Maximization (EM) algorithm by estimating a complex model of health risks, labor market participation, early retirement, and wealth accumulation. I rely on the framework of a dynamic programming discrete choice (DPDC) model that is estimated using data from the German Socio-Economic Panel Study (GSOEP). My analysis focuses on single males and, thus, allows abstracting from adjustments of the partner's labor supply and retirement behavior. The analyzed mechanisms are also relevant for couples and single females. Second, I simulate scenarios where health shocks do or do not occur at different ages during the life-cycle for individuals with differing endowments. A comparison of consumption paths and net present values (NPVs) of expected lifetime consumption between the scenarios sheds light on the health-related risks of consumption and old age poverty that are uninsured by the German social security system. Third, I simulate the introduction of minimum pension benefits that protect individuals from the risk of old age poverty. The simulations indicate only a small decrease in expected retirement age if the minimum benefits are means-tested. The implications of the analysis may also apply to other countries with similar institutions.

DPDC models provide a good framework for the estimation of structural life-cycle models. Under the assumption of revealed preferences, micro data can be used to estimate parameters that characterize the preferences and beliefs of forward looking individuals. Starting with the paper of Wolpin (1984), a literature on structural life-cycle models has evolved that estimates increasingly complex models (see Aguirregabiria and Mira (2010) for an overview). A big advantage of structural models lies in the possibility to perform *ex-ante* counterfactual simulations. However, estimating these kinds of models requires solving a dynamic programming problem that is nested in the estimation criterion. If the state space is large and if unobserved heterogeneity is allowed for, computation can be burdensome. This paper resorts to an extension of the EM algorithm as proposed by Arcidiacono and Jones (2003) for the estimation of finite mixture models with time-constant unobserved heterogeneity. They show how an extended EM algorithm can facilitate the estimation with little loss in efficiency by allowing for a sequential estimation of the parameters (see Arcidiacono (2004,2007) for other applications). I apply the extended EM algorithm to obtain good starting values for a subsequent full information maximum likelihood (FIML) estimation. The efficient FIML estimation procedure only converges with good starting values.

Previous studies use structural models to investigate the link between health and the economic situation of households. Bound et al. (1999) and Disney et al. (2006) show that older workers who are in good health status are more likely to be employed, and Blau and Gilleskie (2008) estimate that bad health halves the employment probability of older employees who have health insurance in the US. Boersch-Supan (2001) explores the incentive effects of the German statutory pension insurance scheme using an option value model. A few studies modeled wealth accumulation within the framework of a life-cycle model. While van der Klaauw and Wolpin (2008) analyze the effects of employment on wealth accumulation, de Nardi et al. (2010) show that health costs explain a large share of households' saving behavior in the US. French and Jones (2011) investigate the effect of life expectancy on optimal saving behavior of retirees. A study by Haan and Prowse (2011) uses a life-cycle model to estimate the effects of an exogenous increase in life expectancy on employment, retirement behavior and savings in Germany. Low and Pistaferri (2011) analyze the insurance value of disability benefits and incentive costs within a life-cycle framework. They find that household welfare can be raised by making program participation less restrictive.

Average pension benefits of early retirees who enter early retirement due to bad health status have declined nominally from 706 EUR to 600 EUR between 2000 and 2010 in Germany, and each year about 160,000 of these early retirees are registered newly with the German statutory pension insurance scheme (Deutsche Rentenversicherung (2011)). Since pension benefits often are the only income source of these early retirees (Albrecht et al. (2007)) and individuals who are in bad health status usually cannot compensate for reductions in the level of pension benefits by delaying retirement, there is a concern of health-related poverty. A quantification of the effects of health shocks on individuals' consumption paths and NPVs of expected lifetime consumption provides important information for policy makers. My simulations suggest that both the risk of old age poverty and health-related changes in these risks depend substantially on an individual's endowments and that the health-related changes may be sizeable. The simulated health-related losses in the NPV of expected lifetime consumption at age 40 that are uninsured by the German social security system are ranging between 3% and 7%. The simulations of the introduction of minimum pension benefits at the poverty line indicate only a small decrease in expected retirement age (between 0 and 0.4 years depending on endowments) if the minimum benefits are means-tested. Overall, means-tested minimum pension benefits appear to be a practicable approach in order to protect individuals from the risk of old age poverty.

The paper is structured as follows. I begin with a presentation of the data, some descriptive statistics, and the institutional framework. Then, I proceed with an outline of the life-cycle model and estimation approach before discussing the parameter estimates. I continue with a presentation of the policy analysis on health-related consumption and poverty risks and the counterfactual reform. A final section summarizes the main findings of the analysis.

4.2 Data and descriptive statistics

My analysis is based on data from the German Socio-Economic Panel study (GSOEP). The GSOEP started in 1984 and collects annual information at both the household and individual levels (Wagner et al. (2007)). I construct an unbalanced panel of single males covering the years 2004 through 2010. The model is estimated for the waves after the 2005 reform of the German income tax system.[39] The survey year 2004 is only included to provide information for the lagged variables of the model. The sample is restricted to single males aged 40 to 64 years with no children in the household. Self-employed, civil servants, and people in institutions are excluded from the sample. The focus on single males allows abstracting from adjustments of the partner's labor supply and retirement behavior. The analyzed mechanisms are also relevant for couple households and single females. I consider the age cohorts 40 to 64 because early retirement is rarely observed in younger cohorts. The final sample consists of 594 independent observations and, in total, of 2,016 observations. Early retirees are only contained in the sample in the first year of retirement because retirement is modeled as an absorbing state and estimation of the model does not require data on individuals after retirement. There are 57 individuals in the sample who opt for early retirement during the observation period. The consumer price index is used to adjust nominal variables to 2005 prices. As a consequence, changes in the wage, net wealth, and saving variables

[39] On January 1, 2005, a reform of the German tax system came into effect lowering the marginal tax rates and making some changes to the tax base. The focus on waves with a relatively homogeneous institutional framework facilitates the computation of the value functions (see below).

can be interpreted as changes in real values. Table 4.1 presents some descriptive statistics on the variables that are used in the analysis. These variables are discussed in more detail in the following paragraphs.

Health status: The GSOEP provides annual information on individuals' health status by both a measure of legally attested disability status and of self-assessed health (SAH). Legally attested disability status is based on a medical examination and has the advantage of being comparatively objective. However, there may be a lag between the realization of a health shock and the completion of the process leading to the approval of the disability status. Furthermore, this measure may not capture some forms of mental illnesses or physical impairments that are below the level required for legally attested disability.[40] These health problems are relevant when investigating the effects of bad health on economic outcomes. For this reason, I combine the objective disability measure with the subjective SAH measure, which presumably captures a broader range of both mental and physical health problems. Furthermore, the SAH measure is found to reflect longitudinal changes in the objective health status reasonably well (Benitez-Silva and Ni (2008)). I construct a binary health measure that defines good health as neither being officially disabled nor assessing own health as "bad" or "very bad". By this definition, about three quarters of the individuals in the sample are in good health status.

Table 4.1: Descriptive statistics

Variable	Mean	Std.	Min	Max
Age	49.5	6.71	40	64
Good health	0.74	0.44	0	1
Employed	0.72	0.45	0	1
Wage	16	5.67	7.52	34.8
Education	12.4	2.46	7	18
Work experience	24.6	8.47	1	48
East Germany	0.3	0.46	0	1
Total savings	3,154	4,243.5	0	33,335
Net wealth	78,310	132577.7	-342,971	1,146,697

Labor market: Since part-time employment is empirically not relevant for the subpopulation of single males, employment behavior is only differentiated between non-employment and full-time employment. Employment is defined as working at least 20 hours per week and the median hours of work for the

[40]The eligibility criteria of the early retirement scheme are independent of the legal disability status.

employees is 40. The share of employed individuals amounts to 72% in the sample. Observed gross wages range from 7.52 to 34.8 EURO per hour, the median wage being 15.2.[41] Education is measured as years of education. The GSOEP constructs the years of education variable from the respondents' information on the obtained level of education and adds some time for additional occupational training. Work experience is defined as years of full-time experience.[42] A binary variable indicates residence in West or East Germany.[43]

Wealth and savings: The GSOEP contains wealth information only for the years 2002 and 2007. Thus, net wealth must be imputed for the 2005, 2006, 2008, 2009, and 2010 waves. This can be done by using information on the individuals' saving behavior. I proceed in a similar way as Fuchs-Schuendeln (2008) defining total savings as the sum of financial and real savings. Since saving information in the GSOEP is left-censored (dissavings are unobserved), I have to assume that individuals aged 40 to 64 do not dissave, but do only decide how much wealth they like to accumulate until retirement. Other studies have derived annual wealth from information on asset income and home ownership (e.g. Fuchs-Schuendeln and Schuendeln (2005), Haan and Prowse (2011)). But, this measure suffers from weaknesses that follow from fluctuations in the returns of individual asset portfolios and from the fact that some assets do not generate any annual returns.

The GSOEP participants indicate their financial savings annually by answering a question about the "usual" amount of monthly savings.[44]. Real savings are defined as annual amortization payments. Since the GSOEP questions ask for the sum of amortization and interest payments, the share of interest payments must be derived from information on the amount of debts. For this purpose, I assume that individuals borrow money at a real interest rate of 6% and receive a real interest rate of 2% on both their financial and real savings. Under these assumptions, debts and net wealth are carried forward from the year 2007 to the other survey years and amortization payments can be differentiated from interest payments.[45]

[41] Some outliers with wages below 7.5 or above 35 EURO are excluded from the sample.

[42] One year of pre-sample part-time experience is counted as half a year of full-time experience.

[43] Berlin is counted as East Germany. By this definition, about 30% of the individuals reside in East Germany. This is slightly more than in the full sample of all individuals where about 27% reside in East Germany.

[44] Question: "Do you usually have an amount of money left over at the end of the month that you can save for larger purchases, emergency expenses or to acquire wealth? If yes, how much?" This measure should be unaffected by seasonal fluctuations.

[45] One observation with net wealth larger than 3,000,000 EURO as well as a few observa-

It turns out that the average saving rate for my sample population (11.7%) is relatively close to the average household saving rate that has been derived by the federal reserve bank from information of the national accounts for the survey years under consideration (11%). Average net wealth is 78,787 EURO, but the median of net wealth is only 23,806 EURO. A share of 25% of the individuals does not have any positive net wealth. There is a share of 31.4% of the individuals who do not save any positive amount.

Survival probabilities: Information on conditional survival probabilities originates from life tables of the Human Mortality Database.[46] I average the age-specific conditional survival probabilities over the years 2005 through 2010, which are used for the estimation of the model.[47]

4.3 Institutional framework

Individuals are making their labor market participation, retirement and saving decisions within the framework of the German tax and transfer system as well as of the German statutory pension insurance scheme. My life-cycle model captures the main features of this institutional framework when computing individuals' net income. For this reason, I outline the main aspects of these institutions insofar as they are relevant for single males during the observation period.

4.3.1 Tax and transfer system

Employed individuals have to pay income tax on both their gross wages and capital income. The income tax rate increases with an individual's taxable income and is payable on all gross income in excess of the income tax allowance. I simplify the analysis by assuming that all individuals receive a real interest rate of 2% on their net wealth after taxes. Furthermore, individuals pay mandatory social security contributions for health, pension, and unemployment insurance. Social security contributions are paid at a constant rate on gross income above a lower limit and below an earnings cap, where half of the contributions are paid by the employer. The contributions of the employees amount to 21.5% of

tions with unrealistically high savings are excluded from the sample.

[46]University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (data downloaded on 5.12.2012).

[47]Using different mortality rates by survey year would require the computation of different values functions for each survey year and increase the computational burden substantially.

the gross wage.

Unemployed individuals are eligible for either unemployment insurance benefits or means-tested social assistance benefits, where the former is proportional to the last net earnings (60% for single households) and the latter ensures a minimum income that does not depend on the individual's employment history. While unemployment insurance benefits are paid for an entitlement period only, social assistance benefits are paid indefinitely.[48] Social assistance benefits comprise a basic amount plus the costs of rent and energy consumption.[49]

4.3.2 Statutory pension insurance scheme

The statutory retirement age of the individuals is 65 years.[50] After retirement, individuals receive public pension benefits.[51] An individual's employment and earnings history is reflected by the pension benefits. The benefits are a deterministic function of the accumulated weighted pension points.[52] An individual accumulates one pension point for each year of employment, which is given a weight of $\frac{wage_{nt}}{\overline{wage_t}}$, where $wage_{nt}$ is individual wage in year t and $\overline{wage_t}$ is the average wage in year t.[53] This leads to complicated dynamic incentives that may be captured by a sufficiently rich life-cycle model. Individuals obtain additional pension points for child care, military service, and during periods of unemployment.[54] I predict a distribution of pension points for the individuals in my sample that is in line with statistics provided by Himmelreicher and Stuchlik (2008) on the basis of administrative data from the German statutory pension insurance scheme.

[48]The entitlement period of unemployment insurance benefits differs by age and employment history. I assume that all the individuals are entitled for one year of unemployment insurance benefits after becoming unemployed. A similar simplification has been made by Adda et al. (2009).

[49]The basic monthly amount is 345 EURO in West and 331 EURO in East Germany. Individuals are assumed to receive 300 EURO for rent and 50 EURO for energy consumption.

[50]Between 2012 and 2029, the statutory pension age will be gradually increased from 65 to 67 years for the cohorts born after 1954. The analysis abstracts from this reform.

[51]Since 2005, pensioners must pay income tax on 50% of their pension benefits. Between the years 2005 and 2040, the share of pension benefits that is subject to income taxation will gradually increase up to 100%.

[52]In 2005, the value of one pension point amounted to monthly benefits of 26.13 EURO in West and 22.97 EURO in East Germany.

[53]There is a year-specific cap on pension point weights, which has been approximately 2 during the observation period.

[54]Individuals who receive unemployment insurance benefits also pay social security contributions and receive 0.8 pension points that are weighted according to the last wage. Pension points for individuals who receive social assistance benefits are negligible. I assume that individuals have received unemployment benefits during up to two years of their periods of non-employment when computing the number of pension points.

The possibility of early retirement in the German statutory pension insurance scheme constitutes an insurance against work incapacity. Individuals may opt for early retirement if they are eligible. Eligibility depends on age, employment history, and health status. Individuals with a sufficiently long employment history (more than 35 years of work) are eligible for early retirement if aged 63 or over. Unemployed individuals are eligible if aged 60 or over.[55] Boersch-Supan (2001) points out that unemployment as a transition to early retirement is likely to be endogenous and to be a strategic variable of both the employer and employee. Individuals who are in sufficiently poor health status (work incapacity) can opt for early retirement even before the age of 60. If these individuals retire before the age of 60, they can claim pension benefits as if they had accumulated pension points until age 60. Individuals who want to opt for early retirement because of bad health must pass a medical examination that is performed by a physician from the statutory pension insurance scheme. While it is hard to believe that individuals who actually are in good health status can easily pass the examination, it appears to be likely that individuals who are in bad health status, but are not work incapacitated, can pass by exaggerating their health problems. For each year of early retirement before the age of 65, 63 for individuals in bad health status, a penalty of 3.6% is applied on the pension benefits. This penalty is only applied up to a maximum of 18%, 10.08% for individuals in bad health status. Hence, bad health opens up the option of early retirement before the age of 60 and goes along with lower reductions on the pension benefits of early retirees.

4.4 Model and specification

In the following subsections, I outline the structural life-cycle model in greater detail. After summing up its main features, I proceed with a presentation of the various components of the model.

4.4.1 Main features

Framework: The methodological framework is similar to Rust (1987) and Rust and Phelan (1997), but accounts for wealth accumulation and time-constant unobserved heterogeneity. Individuals maximize their expected lifetime utility by making decisions about labor market participation, early re-

[55] Between 2006 and 2010, the minimum age for early retirement because of unemployment has increased gradually from 60 to 63.

tirement and saving behavior in each period of time (annual data). The set of possible choices is restricted by eligibility requirements for early retirement and by job offer and separation rates that are estimated within the model. The job offer and separation rates are estimated differentially by health status. Individuals who do not opt for early retirement retire upon reaching the statutory pension age of 65 years. Individuals have rational expectations and are facing a dynamic programming problem with a finite horizon. They have preferences about consumption and leisure time. Unlike other studies that resort to a two-step estimation approach (e.g. French (2005), French and Jones (2011)), the model allows for correlations between the unobserved heterogeneity in the leisure preferences and the unobserved components in both the health process and wage equation. This accounts for selection into the labor market.

Health: Health is modeled as an autoregressive process that depends on lagged health status, years of education, and age. Since it is not observed whether a health shock actually leads to a work incapacity, the model assumes that individuals who are in bad health status can work, but that their labor market risks (captured through job offer and separation rates), the eligibility requirements for early retirement, and the financial incentives of the early retirement scheme explain the effect of bad health on the probability of labor market participation and early retirement. Since job offer and separation rates capture persistence of the unemployment status that is also taken into account by forward looking individuals, unemployed individuals who are in bad health status may have a strong incentive to opt for early retirement. Given the presumably low job offer rates of these individuals, their opportunity costs of early retirement are comparatively low.

Budget: Wages are estimated within the model (as a function of years of education, work experience, and region of residence) and pension benefits are a deterministic function of retirement age, work experience, wage history, and some other relevant factors. This leads to complicated dynamic incentives that are captured by the DPDC framework. Individuals take into account both the effect of the human capital accumulation process on wages (Eckstein and Wolpin (1989)) and the respective effects on future pension benefits when deciding about labor market participation and early retirement. When computing the individuals' net income, I apply the rules and regulations of the German tax and transfer system and of the German statutory pension scheme. After retirement, individuals dissave according to the value of an actuarially fair life annuity that could be bought with the accumulated wealth.

The model includes the following **state variables**: individual's age, net wealth, work experience, years of education, residence in East or West Germany, health status, labor supply restriction and previous period's choices. The **choice problem** consists of two parts: of an optimal stopping problem regarding the early retirement decision and of labor market participation and saving choices before retirement.

4.4.2 Objective function

I specify a DPDC model of individuals' labor market participation, early retirement and saving decision. Individuals are finitely lived and die no later than period T, which is set to be 100. Discrete time is indexed by t (individual's age), and there is a number of N individuals, indexed by n. Each individual n receives a utility flow $U(\mathbf{s}_{nt}, d_{nt})$ in period t where $\mathbf{s}_{nt}$ is a vector of state variables, and d_{nt} indicates the individual's choice. Note that the saving decision is discretized. An individual chooses between nine alternative choices: working and zero, [500,1500), [1500,5000), [5000,10000), or [10000,∞) EURO of annual savings; not working and zero, [500,1500), [1500,∞) EURO of annual savings; early retirement (if eligible) and dissaving according to the value of an actuarially fair life annuity that could be bought with the accumulated wealth. Individuals save a non-negative amount before retirement.[56] Individuals are assigned the median savings of the respective saving category and retire no later than the statutory pension age of 65 years. Every period t, an individual observes the state variables $\mathbf{s}_{nt}$ and makes the decision d_{nt} that maximizes expected lifetime utility:

$$E\left\{\sum_{j=0}^{T-t} p_{t+j}\beta^j U(\mathbf{s}_{nt+j}, d_{nt+j})\right\} \tag{4.1}$$

where β is the time discount factor, which is set to be 0.96 (Gourinchas and Parker (2002) provide a reliable estimate) and p_{t+j} is the conditional survival probability of the individual for period t+j given survival until period t.[57]

[56]This assumption is due to a data restriction of the saving information (left-censored). An exception are unemployed individuals who are not eligible for unemployment insurance benefits and fail the means test required for social assistance benefits. These individuals receive an income at the minimum income level that is deducted from their net wealth, where 10,000 EURO are exempted from the means test that is required for social assistance benefits. The exemption level of 10,000 EURO is assumed because the actual rules are very complicated and enforcement of these rules is unobserved.

[57]Information on conditional survival probabilities originates from life tables of the Human Mortality Database.

4.4.3 Utility function

Individuals have preferences about consumption and leisure time that are represented by the following time separable random utility model:

$$U(\mathbf{s}_{nt}, d_{nt}) = \alpha_1 \frac{c(\mathbf{s}_{nt}, d_{nt})^{(1-\rho)} - 1}{(1-\rho)} + \alpha_{2n}(1 - work(d_{nt})) + \epsilon_{nt}(d_{nt}) \tag{4.2}$$

where $\epsilon_{nt}(d_{nt})$ is assumed to be type 1 extreme value distributed. $c(\mathbf{s}_{nt}, d_{nt})$ is the level of consumption that is associated with state $\mathbf{s}_{nt}$ and decision d_{nt}. $work(d_{nt})$ indicates employment such that $(1 - work(d_{nt}))$ captures the leisure time that is associated with either non-employment or retirement. α_1 is a consumption weight and ρ is the coefficient of relative risk aversion (CRRA). Unobserved heterogeneity in the leisure preferences is reflected by α_{2n}. The parameters of the utility function are contained by the vector $\boldsymbol{\theta}_U = (\alpha_1, \rho, \alpha_{2n})$.

4.4.4 Value function

The individuals' beliefs about future states are captured by a Markov transition function $q(\mathbf{s}_{nt+1}|s_{nt}, d_{nt})$ that is indicating the respective transition probabilities. In particular, $q(\mathbf{s}_{nt+1}|s_{nt}, d_{nt})$ captures the individual expectations about the transitions of the health status that is evolving stochastically over time. Furthermore, it takes into account the expectations of unemployed individuals to receive a job offer and of employed individuals to face a job separation in the following period (see next subsection). For state variables like net wealth or work experience that evolve deterministically, the probability of the determined state is one while it is zero for all other possible states of the variable. Since there is a discrete set of possible future states, $q(\mathbf{s}_{nt+1}|\mathbf{s}_{nt}, d_{nt})$ is a probability mass function and not a density function. By Bellman's principle of optimality, the value function $V_t(\mathbf{s}_{nt})$ can be represented recursively as

$$V_t(\mathbf{s}_{nt}) = \max_{d_{nt} \in D(\mathbf{s}_{nt})} \left\{ U(\mathbf{s}_{nt}, d_{nt}) + p_{t+1}\beta \int_{\epsilon} \left[\sum_{\mathbf{s}_{nt+1}} V_{t+1}(\mathbf{s}_{nt+1}) q(\mathbf{s}_{nt+1}|\mathbf{s}_{nt}, d_{nt}) \right] g(\epsilon_{nt+1}) \right\} \tag{4.3}$$

where $D(\mathbf{s}_{nt})$ is the choice set that is available to individual n in period t and $g(\cdot)$ is the probability density function of the unobserved random components of the utility function. Individuals only make choices until they reach the statutory pension age of 65 years and the available choice set depends on the individual's state, $\mathbf{s}_{nt}$. The choice set is restricted by the eligibility requirements for early retirement (see section 3) and by job offer and separation rates that are estimated within the model. Since it is unobserved whether a health shock actually leads to a work incapacity, the model makes the assumption that individuals who are in bad health status are still capable of working, but that their labor market risks (captured through job offer and separation rates), the eligibility requirements for early retirement, and the financial incentives of the early retirement scheme (lower penalty on pension benefits and pension points as if the individuals had worked until age 60) explain the effect of bad health on the probability of labor market participation and early retirement.

4.4.5 Job offer and separation rates

An individual's choice of labor market participation is restricted by job offer and separation rates that are estimated within the model. This captures persistence in the employment status. Individuals who have been unemployed in the previous period may only choose labor market participation if they receive a job offer in the current period. Analogously, individuals who have been employed in the previous period may only choose labor market participation if they do not face a job separation in the current period. The job offer and separation rates are estimated differentially by health status in order to account for the change in labor market risks that is induced by a health shock. If the choice of labor market participation is restricted, individuals can only choose between unemployment and early retirement (if eligible). It follows from the persistence of the unemployment status - being taken into account by forward looking individuals - that unemployed individuals may have a strong incentive to opt for early retirement (in particular when health status is bad). Individuals receive a job offer with a probability of either ϕ_{gh}^{offer} or ϕ_{bh}^{offer} and face a job separation with a probability of either ϕ_{gh}^{sep} or ϕ_{bh}^{sep}, depending on their health status. The vector $\boldsymbol{\phi} = (\phi_{gh}^{offer}, \phi_{bh}^{offer}, \phi_{gh}^{sep}, \phi_{bh}^{sep})$ contains the job offer and separation rates.

4.4.6 Health process

In order to take into account the high observed persistence in the individuals' health status, good health is modeled as an autoregressive process that depends on lagged health status, education, and age. Education captures differences in the health risks by socio-economic status.[58] The probability of good health is given by

$$P(health_{nt} = 1) = \Lambda(\psi_1 health_{nt-1} + \psi_2 educ_{nt} + \psi_3 age50_{nt} + \tau_n) \qquad (4.4)$$

where $\Lambda(\cdot)$ is the logistic distribution function, $health_{nt}$ indicates good health status, $educ_{nt}$ is years of education, and $age50_{nt}$ indicates an age of 50 years or older. The inclusion of $age50_{nt}$ as an explanatory variable is in line with the observed health pattern in the data. Unobserved time-constant heterogeneity is captured by τ_n. The vector $\boldsymbol{\theta}_h = (\psi_1, \psi_2, \psi_3, \tau_n)$ contains the parameters of the health process.

4.4.7 Gross wage

Gross wages are assumed to follow a log-normal distribution.[59] The logarithm of gross wages is modeled as

$$\log(wage_{nt}) = \delta_1 educ_n + \delta_2 ex_{nt} + \delta_3 east_{nt} + \kappa_n + \mu_{nt} \qquad (4.5)$$

where $educ_n$ is years of education, ex_{nt} is years of work experience, $east_{nt}$ is a dummy variable for residence in East Germany, κ_n is time-constant unobserved heterogeneity, and μ_{nt} is i.i.d. $N(0, \sigma_\mu)$. It follows from the DPDC framework that individuals are taking into account the human capital accumulation process when making their labor market participation decision (Eckstein and Wolpin (1989)). Hence, work experience is an endogenous variable in the model. The wage equation does not include $health_{nt}$ as an explanatory variable because it is empirically not relevant.[60] The correlation between the individual-specific leisure preferences in the utility function and the unobserved component, κ_n, in the wage equation accounts for selection into the labor market. The vector $\boldsymbol{\theta}_w = (\delta_1, \delta_2, \delta_3, \kappa_n, \sigma_\mu)$ contains the parameters of the wage

[58] Quasi-experimental evidence suggests that education even exerts a causal effect on health and health behavior (see e.g. Kemptner et al. (2011) for Germany).

[59] Individuals are assumed to receive the expected wage if they participate in the labor market. From a theoretical point of view, this is justified if the transitory component of observed wages is either attributed to measurement error or if individuals only observe it after the labor market participation decision has been made. When computing gross labor earnings, I assume that individuals work the median number of hours, which is 40 in the sample.

[60] When estimating a specification with $\log(wage_{nt})$ as a function of health status, the respective coefficient estimate is small and insignificant. It appears that individuals who are in bad health status either do not work or receive a wage according to their qualification. Of course, the qualification in terms of work experience may depend on past health outcomes. This is taken into account by the model.

equation.

4.4.8 Budget constraint

Individuals face a budget constraint when making their saving/consumption decision which comprises three equations:

$$\begin{aligned} c(\mathbf{s}_{nt}, d_{nt}) &= G\left(\mathbf{s}_{nt}, d_{nt}\right) - savings(d_{nt}) \\ wealth_{nt+1} &= (1+r_t)\left(wealth_{nt} + savings(d_{nt})\right) \\ wealth_{nt} &> 0 \end{aligned} \tag{4.6}$$

where $c(\mathbf{s}_{nt}, d_{nt})$ is the level of consumption associated with state $\mathbf{s}_{nt}$ and decision d_{nt}, and $G(\cdot)$ indicates net income by applying the rules and regulations of the German tax and transfer system and of the statutory pension insurance scheme.[61] I assume that the forward looking individuals neither expect future changes in the institutional framework nor do they expect a change in their economic situation that is due to finding a partner. $wealth_{nt}$ is period t's net wealth, r_t is the real interest rate that is set to be 0.02 after taxes, and $savings(d_{nt})$ is the amount of savings being associated with state $\mathbf{s}_{nt}$ and decision d_{nt}. I assume that retirees dissave according to the value of an actuarially fair life annuity that could be bought with the accumulated wealth. A more detailed modeling of retirees' saving behavior is not necessary for my research questions. The first equation defines the possible levels of consumption in period t, the second equation describes the wealth accumulation process, and the third equation is a non-negativity constraint being a standard assumption in the life-cycle literature.

4.4.9 Unobserved heterogeneity and initial conditions

Following the approach of Heckman and Singer (1984), unobserved heterogeneity is accounted for semi-nonparametrically by allowing for a finite number of unobserved types $m \in 1, ..., M$. Each type comprises a fixed proportion of the individuals in the population. Hence, the individual-specific parameters α_{2n}, τ_n, and κ_n are assumed to be equal to the respective type-specific parameters

[61] Since only marginal changes to the institutional framework between 2005 and 2010 occurred, I apply the rules and regulations from the year 2005 to all survey years. This simplifies the estimation of the model and saves computational time because I do not have to estimate a different value function for each survey year.

α_{2m}, τ_m, and κ_m. The probability that individual n is of type m is given by

$$\pi_{mn} = \frac{exp(\boldsymbol{\gamma}_m \mathbf{z}_n)}{1 + \sum_{l=1}^{M-1} exp(\boldsymbol{\gamma}_l z_n)}, \; for \; m = 1, ..., M-1 \tag{4.7}$$

where $\boldsymbol{\gamma}_M$ is normalized to zero and $\sum_{m=1}^{M} \pi_{mn} = 1$. The vector $\mathbf{z}_n$ contains initial observed health and employment status, and the interaction terms between a variable indicating an age of 50 years or older in the first wave that includes the individual and initial observed health and employment status. The interaction terms account for individuals entering the sample at different ages. I deal with the initial condition problem (Heckman (1981b)) by assuming that the unobserved true initial conditions (the process starts before the age of 40) are exogenous conditional on type. This follows the idea of Wooldridge (2005).[62] As opposed to an approach that assumes that the structural model can explain the distribution of the initial values of the state variables, this is computationally much less intense and does not rely on potentially unrealistic out-of-sample extrapolations (Aguirregabiria and Mira (2010)).

4.4.10 Choice probabilities and log-likelihood

Given the finite horizon of the individual's optimization problem, it can be solved recursively. The expected value function, $v_t(\mathbf{s}_{nt}, d_{nt})$, for period T is simply given by this period's expected utility flow:

$$v_T(\mathbf{s}_{nT}, d_{nT}) = u(\mathbf{s}_{nT}, d_{nT}) \tag{4.8}$$

By Bellman's principle of optimality, the individual's optimization problem can be written as a two-period problem for the other periods of time. It follows from the type 1 extreme value distribution of $\epsilon_{nt}(d_{nt})$ that the expected value function has a closed form solution (Rust (1987)):

$$\begin{aligned} v_t(\mathbf{s}_{nt}, d_{nt}) =& u(\mathbf{s}_{nt}, d_{nt}) + p_{t+1}\beta \\ & \sum_{\mathbf{s}_{nt+1}} \log \left\{ \sum_{d_{nt+1} \in D(\mathbf{s}_{nt+1})} exp(v_{t+1}(\mathbf{s}_{nt+1}, d_{nt+1})) \right\} q(\mathbf{s}_{nt+1} | \mathbf{s}_{nt}, d_{nt}) \end{aligned} \tag{4.9}$$

The computation of the expected value functions for periods t=65,...,T-1 is comparatively simple because individual choices are only modeled for t=40,...,64.

[62] Akay (2011) performs various Monte Carlo experiments showing that such an approach works reasonably well for panel data sets of moderately long duration.

Rust (1987) shows that under the assumptions of additive separability and conditional independence, the conditional choice probabilities have a closed form solution (here: mixed logit probabilities):

$$Prob(d_{nt}|\mathbf{s}_{nt}) = \frac{exp(v_t(\mathbf{s}_{nt}, d_{nt}))}{\sum_{j \in D(\mathbf{s}_{nt})} exp(v_t(\mathbf{s}_{nt}, j))} \tag{4.10}$$

When computing choice probabilities, I take into account that the choice of labor market participation is restricted with a probability of 1-$\boldsymbol{\phi}_{gh}^{offer}$ or 1-$\boldsymbol{\phi}_{bh}^{offer}$ for individuals who did not work in the previous period and with a probability of $\boldsymbol{\phi}_{gh}^{sep}$ or $\boldsymbol{\phi}_{bh}^{sep}$ for individuals who did work in the previous period, where the respective rates differ by an individual's health status.

The expected value functions are only computed for a discretized state space in order to save computational time (Keane and Wolpin (1994)). As a consequence, interpolation methods must be used to approximate the functions at the observed values of the state variables. I resort to a cubic spline function - as recommended e.g. by Adda and Cooper (2003) for the estimation of dynamic models - in order to interpolate the expected value functions for net wealth, work experience, and years of education. For each of these variables, I define five grid points. My results are insensitive to an increase in the number of grid points.

The log-likelihood function of the sample is given by

$$\sum_{n=1}^{N} \log \left\{ \sum_{m=1}^{M} \pi_{mn}(\boldsymbol{\gamma}) \prod_{t=1}^{T} L_m(d_{nt}|\boldsymbol{\theta}_U, \boldsymbol{\phi}, \boldsymbol{\theta}_h, \boldsymbol{\theta}_w) L_m(health_{nt}|\boldsymbol{\theta}_h) L_m(wage_{nt}|\boldsymbol{\theta}_w) \right\} \tag{4.11}$$

where $L_m(d_{nt}|\boldsymbol{\theta}_U, \boldsymbol{\phi}, \boldsymbol{\theta}_h, \boldsymbol{\theta}_w)$ is the likelihood contribution of the observed decision d_{nt} of individual n in period t, if n is of type m. The likelihood contributions of the health process and wage equation are given by $L_m(health_{nt}|\boldsymbol{\theta}_h)$ and $L_m(wage_{nt}|\boldsymbol{\theta}_w)$, respectively.

4.5 EM algorithm and FIML

In principle, the model can be estimated by finding the maximum of the log-likelihood function. However, it is due to the non-separability of the log-likelihood function that a stepwise maximization is impossible. A direct maximization with respect to all parameters involves considerable numerical problems and is computationally very intense. For this reason, I resort to an extension of the EM algorithm as proposed by Arcidiacono and Jones (2003) for the estimation of finite mixture models with time-constant unobserved heterogeneity. They show how an extended EM algorithm can be used to facilitate the estimation with little loss in efficiency by allowing for a sequential estimation of the parameters.

By Bayes rule, the conditional probability Π_{mn} that individual n is of type m, given the observed choices as well as $\boldsymbol{\theta}_U, \boldsymbol{\phi}, \boldsymbol{\theta}_h, \boldsymbol{\theta}_w$, and $\boldsymbol{\gamma}$, can be written as

$$\Pi_{mn} = \frac{\pi_{mn}(\boldsymbol{\gamma}) \prod_{t=1}^{T} L_m(d_{nt}|\boldsymbol{\theta}_U, \boldsymbol{\phi}, \boldsymbol{\theta}_h, \boldsymbol{\theta}_w) L_m(health_{nt}|\boldsymbol{\theta}_h) L_m(wage_{nt}|\boldsymbol{\theta}_w)}{\sum_{m=1}^{M} \pi_{mn}(\boldsymbol{\gamma}) \prod_{t=1}^{T} L_m(d_{nt}|\boldsymbol{\theta}_U, \boldsymbol{\phi}, \boldsymbol{\theta}_h, \boldsymbol{\theta}_w) L_m(health_{nt}|\boldsymbol{\theta}_h) L_m(wage_{nt}|\boldsymbol{\theta}_w)} \tag{4.12}$$

Using the conditional type probabilities, the following additively separable expected log-likelihood function can be derived:

$$\begin{aligned}\sum_{n=1}^{N} \sum_{m=1}^{M} \sum_{t=1}^{T} \Pi_{mn}(\log(L_m(d_{nt}|\boldsymbol{\theta}_U, \boldsymbol{\phi}, \boldsymbol{\theta}_h, \boldsymbol{\theta}_w))+ \\ \log(L_m(health_{nt}|\boldsymbol{\theta}_h)) + \log(L_m(wage_{nt}|\boldsymbol{\theta}_w)))\end{aligned} \tag{4.13}$$

The EM algorithm reintroduces additive separability at the maximization step. Starting with arbitrary initial values, the maximum of the log-likelihood function can be found by maximizing iteratively the expected log-likelihood function, then using the estimates of $\boldsymbol{\theta}_U$, $\boldsymbol{\phi}$, $\boldsymbol{\theta}_h$, and $\boldsymbol{\theta}_w$ to estimate $\boldsymbol{\gamma}$ by maximizing the log-likelihood function conditionally on these parameter estimates, and finally using all the estimated parameters for updating the posterior type probabilities. Subsequently, the expected log-likelihood function is maximized again. Iterating on these steps until convergence yields the maximum of the log-likelihood function (see Boyles (1983) and Wu (1983) for formal proofs).

The additive separability of the expected log-likelihood function allows maximizing it sequentially. This is done by first estimating $\boldsymbol{\theta}_h$ and $\boldsymbol{\theta}_w$, and then taking these estimates as given in a maximization of the expected log-likelihood function with respect to $\boldsymbol{\theta}_U$. Arcidiacono and Jones (2003) show that such an extension of the EM algorithm produces consistent parameter estimates. While the approach reduces the computational burden, the parameter estimates are inefficient and the estimation of standard errors is complicated because the computational time makes the use of bootstrapping methods unpractical. I apply the sequential EM algorithm in order to obtain good starting values for a subsequent FIML estimation. This also facilitates the computation of standard errors that can simply be derived from the inverse of the Hessian of the log-likelihood function at its maximum.

4.6 Estimation results and model fit

4.6.1 Parameter estimates

The extended EM algorithm gets close to convergence after a number of iterations that depends on the choice of starting values.[63] It has been a convenient approach to abort the algorithm at this point and to use the current trail values of the parameters as initial values in a FIML estimation.[64] While the extended EM algorithm slows down when approaching convergence, the optimization algorithm that is used for the FIML estimation converges comparatively quickly when using good starting values.[65] It is difficult to find the maximum of the log-likelihood function by the FIML approach without good starting values because the optimization algorithm often stops before reaching the maximum. This even happens when raising the stepsize tolerance of the algorithm. The model is estimated allowing for two unobserved types (M=2).[66] I did not achieve convergence of the algorithm for more than two unobserved types. Table 4.2 shows the parameter estimates, the trail values of the extended EM algorithm after 10 iterations, and the starting values used for the extended EM algorithm. In the following, I shortly discuss the estimation results:

Utility function: The estimate of the consumption weight α_1 is reasonable and the coefficient of relative risk aversion, ρ, is estimated to be 1.17. This estimate of ρ is larger than the estimate of Rust and Phelan (1997), but smaller than the estimates of French (2005) and French and Jones (2011). Individuals exhibit unobserved heterogeneity in their leisure preferences that is captured by the type-specific parameter, α_{2m}. The two parameter estimates for the value of leisure time are positive, but significant only for inividuals of type 2.

Job offer and separation rates: The low estimated job offer rates indicate a high persistence of unemployment. This matches the small number of observed transitions from unemployment to employment in the data. Both the job offer and separation rates differ substantially by health status. A health shock is estimated to increase the labor market risks both by lowering the probability of a job offer and by raising the probability of a job separation.

[63] The starting values are shown in table 4.2. Most starting values are set to be 0.1. The estimation results do not hinge on the choice of starting values.

[64] I abort the extended EM algorithm after 10 iterations.

[65] The log-likelihood at its maximum and at the trail values from the EM algorithm after 10 iterations differ only slightly (see table 4.2).

[66] Estimation takes about 11 hours on a standard laptop (i5-2430, 8gb ram, matlab 32 bit): about 8 hours for the 10 iterations of the EM algorithm and about 3 hours for the FIML estimation.

Table 4.2: Parameter estimates

	FIML		EM alg. (10 iterations)	Starting values
	Estimates	St.e.	Trail values	
Utility function:				
α_1 (consumption)	2.727	(0.2400)	2.619	0.1
ρ (crra)	1.165	(0.0271)	1.163	0.1
α_{21} (leisure, type 1)	0.326	(0.3504)	0.509	0.1
α_{22} (leisure, type 2)	1.388	(0.0892)	1.357	0.1
Job offer and separation rates:				
ϕ_{bh}^{sep} (separation, bad health)	0.100	(0.0216)	0.099	0.5
ϕ_{gh}^{sep} (separation, good health)	0.018	(0.0057)	0.018	0.5
ϕ_{bh}^{offer} (offer, bad health)	0.010	(0.0052)	0.010	0.5
ϕ_{gh}^{offer} (offer, good health)	0.074	(0.0145)	0.068	0.5
Wage equation:				
κ_1 (constant, type 1)	2.070	(0.0278)	2.075	2.5
κ_2 (constant, type 2)	1.633	(0.0225)	1.646	1.5
δ_1 (years of education / 10)	0.629	(0.0259)	0.608	0.5
δ_2 (work experience / 10)	0.053	(0.0074)	0.060	0.1
δ_3 (East Germany)	-0.308	(0.0106)	-0.306	0.1
σ_μ (standard deviation)	0.209	(0.0042)	0.209	0.1
Health process:				
τ_1 (constant, type 1)	-1.652	(0.4178)	-1.536	-1.5
τ_2 (constant, type 2)	-1.939	(0.3815)	-1.869	-1.5
ψ_1 ($health_{t-1}$)	3.368	(0.1399)	3.312	1.5
ψ_2 (years of education / 10)	0.739	(0.3010)	0.718	0.1
ψ_3 (age>=50)	-0.604	(0.1383)	-0.532	0.1
Prob. of type 1 (38.4 %):				
γ_0 (constant)	-1.651	(0.6745)	-1.829	0.1
γ_1 (initial health)	0.182	(0.4252)	0.219	0.1
γ_2 (initial health×(age>=50))	-0.610	(0.5997)	-0.579	0.1
γ_3 (initial empl.)	1.437	(0.7235)	1.608	0.1
γ_4 (initial empl.×(age>=50))	0.012	(0.5273)	-0.038	0.1
Estimation time	≈3 hours		≈8 hours	-
Log-likelihood	-4,237.7		-4,239.7	-11,977

Note: The indicated starting values are used for the EM algorithm, which is aborted after 10 iterations. Then, the current trial values of the parameters are used as starting values for the FIML estimation procedure. At last, the standard errors are derived from the inverse of the Hessian of the log-likelihood function at its maximum.

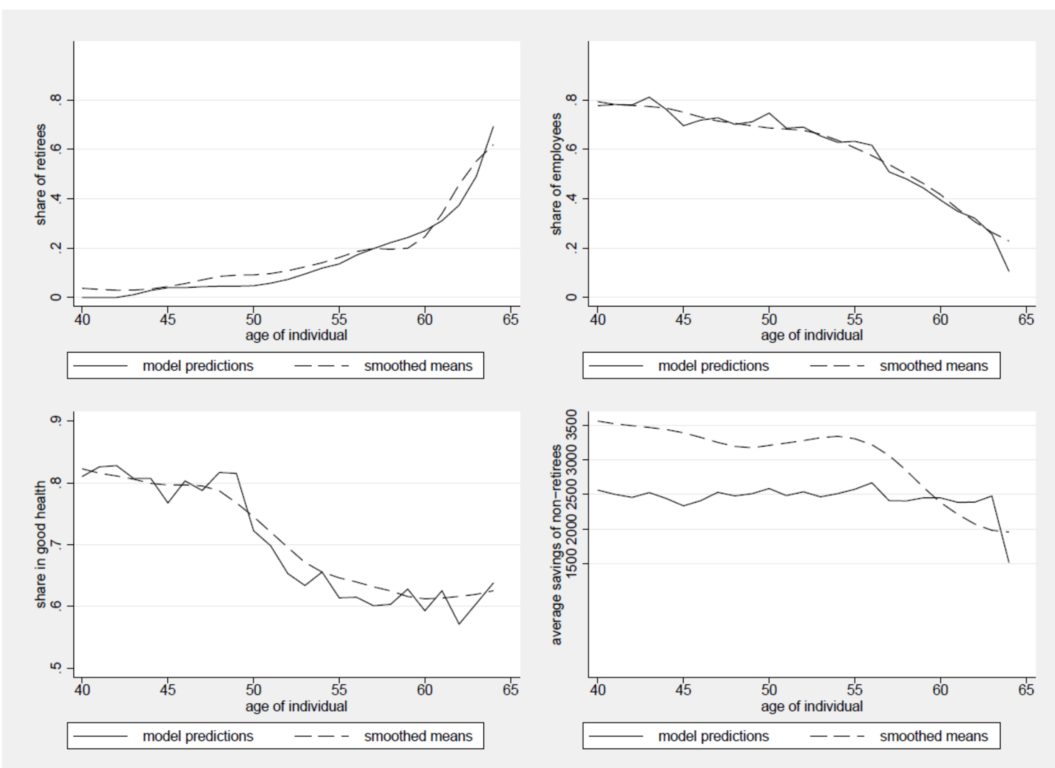
Wage equation: There is substantial unobserved heterogeneity in the productivity of individuals that is captured by the type-specific constant κ_m and that is negatively correlated with the leisure preferences. As expected, years of education and work experience increase wages while residence in East Germany is associated with a lower wage. The latter finding reflects persistent differences in the level of wages between East and West Germany. The standard deviation of the normally distributed μ is estimated to be 0.21.

Health process: The coefficient on lagged health status, ψ_1, points to considerable state dependence. Hence, individuals who experience a health shock tend to remain in bad health status. Years of education are associated positively with an individual's probability of good health. In line with the observed health pattern, an age of 50 years or older decreases the probability of good health.

Type probabilities: The two unobserved types are predicted to comprise 38.4% and 61.6% of the population. Only the estimates for the constant and the coefficient on the initial observed employment status are significant, while the coefficient estimates on the other initial observed conditions are reasonable, but insignificant. The estimates suggest that individuals of type 1 are more likely than individuals of type 2 to enter the sample in employment.

4.6.2 Model fit

Figure 4.1: Comparison of predictions from the life-cycle model and from non-parametric local constant estimations (smoothed means)



I check the model fit by making predictions for the probabilities of early retirement, labor market participation, health status, and saving behavior for all the individuals in my sample. I compare average predictions of the model by age with non-parametrically smoothed means. The smoothed means are computed on the basis of non-parametric local constant estimations of the association between the respective outcomes and age.[67] When computing average predictions for the retirement and employment decisions, I take into account that retirement is modeled as an absorbing state and adjust the predictions such that they refer to the full population of retirees and non-retirees.[68] Figure 4.1 presents the model's predictions and the smoothed means. It turns out that the model produces reasonably good predictions. However, I underpredict savings, which is a common problem of structural models.[69] This is driven by the predicted choice probabilities for the higher saving categories and becomes relevant only for individuals with comparatively high income. Better predictions of the saving behavior presumably require the estimation of heterogeneous risk preferences.[70]

4.7 Policy Analysis

This section investigates the health-related risks of consumption and old age poverty that are uninsured by the German social security system. While the first subsection focuses on the health-related risks within the current institutional framework, the second subsection presents the simulations for a counterfactual reform that introduces means-tested minimum pension benefits.

[67] The local constant estimators rely on a plugin estimator of the asymptotically optimal constant bandwidth (see Fan and Gijbels (1996), StataCorp (2009)) and an Epanechnikov kernel.

[68] I compute the probability of labor market participation as the probability of not having opted for early retirement in previous periods and choosing labor market participation in the current period. The probability of being retired is the probability of either having opted for early retirement in one of the previous periods or opting for early retirement in the current period. The non-parametric estimations for these outcomes must be computed for the full sample of retirees and non-retirees in order to be comparable with the model's predictions. Hence, this sample includes retirees not only in their first year of retirement.

[69] See e.g. van der Klaauw and Wolpin (2008), table 4.

[70] I did not achieve convergence of the optimization algorithm when allowing for additional type-specific coefficients for the coefficient of relative risk aversion, ρ, and the consumption weight, α_1, which then also operates as a scaling factor.

4.7.1 Health-related risks

I make life-cycle predictions for individuals by differing endowments at age 40. Endowments differ with respect to net wealth, work experience, and years of education. Individuals are assumed to be in good health status and to be employed in the year before they turn 40. The simulations are implemented under the assumption that the individuals do not have additional private disability insurance. I simulate five scenarios where health shocks do or do not occur at different ages during the life-cycle. A comparison of consumption paths and net present values of expected lifetime consumption between the scenarios sheds light on the health-related consumption and poverty risks that are uninsured by the German social security system. In the first scenario, no health shock occurs (reference scenario). The second scenario follows the specification of the life-cycle model such that health evolves stochastically according to the estimated health process. In the other three scenarios, a persistent health shock is assumed to occur at the ages 60, 55, and 45, respectively, while no health shock occurs before these ages. A similar approach has been adopted by Haan and Myck (2009) when investigating health and labor market risks. In all scenarios, individuals behave as if health was evolving stochastically.

Expected consumption paths are computed by simulating 5,000 life-cycle paths for each of the varying endowments at age 40 and by taking averages of the simulated consumption paths. I simulate choices and transitions of the health status (for the second scenario) by taking quasi-random draws from the uniform distribution. Furthermore, at the start of each of the simulations, the type is determined by taking a draw from the uniform distribution.[71] All state variables are carried forward between the periods. Analogously to the interpolation of the value function in the estimation of the model, I compute choice probabilities as well as income and consumption functions for a discretized state space and, then, resort to a cubic spline function to interpolate these functions at the simulated values of the state variables.

I consider types of individuals in East or West Germany who are endowed at age 40 with a net wealth (NW) of either zero or 20,000 EURO, having either an employment history without any gaps (No Gap) or having experienced a 5-year-period of non-employment (5Y-Gap), and having completed either 9, 13

[71] The distribution of types corresponds to the estimated distribution of types in the sample population.

or 18 years of education (YE).[72] Investigating such a grid of endowments gives an idea of the magnitude of the risks and shows how these risks interrelate with the endowments.

Lifetime consumption

Figure 4.2: Expected consumption paths for differing endowments at age 40

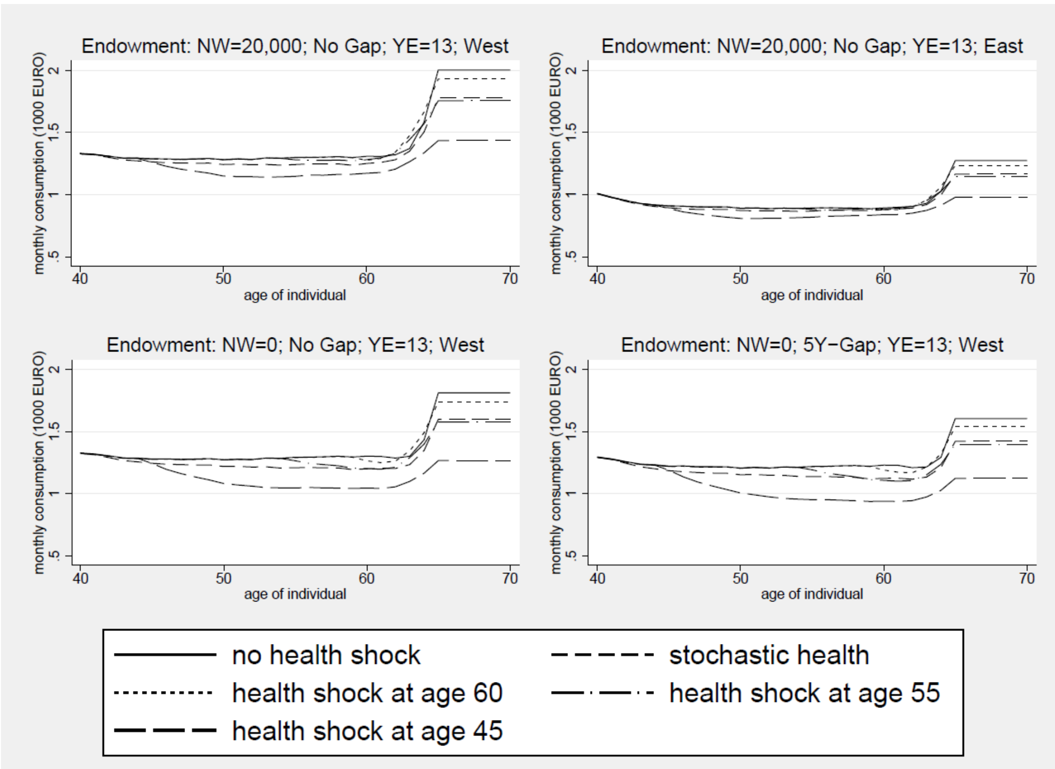


Figure 4.2 presents expected consumption paths for a selection of endowments at age 40 for individuals with a medium level of education. This graph is supposed to give an idea about how health affects expected consumption paths. The increase in expected consumption after retirement follows from the assumption that individuals dissave according to the value of an actuarially fair life annuity that could be bought with the accumulated wealth. This is potential consumption rather than actual consumption (e.g. individuals in

[72]While 9 years correspond to the compulsory schooling, 13 years can be interpreted as either the completion of 10 years of schooling plus 3 more years of vocational training or the completion of an academic track school without any further professional training. 18 years usually correspond to 13 years of schooling in an academic track school plus 5 years of university education.

Table 4.3: Simulated loss of expected lifetime consumption at age 40

	Endowments at age 40			Losses in NPVs			
	Net wealth	5-year gap in empl.	Years of education	Stochastic health	Shock at age 60	Shock at age 55	Shock at age 45
West	20,000	no	9	5.9%	1.1%	3.9%	12.0%
West	20,000	yes	9	6.0%	1.2%	4.0%	13.0%
West	0	no	9	6.3%	1.2%	4.4%	13.5%
West	0	yes	9	6.3%	1.3%	4.5%	14.0%
West	20,000	no	13	6.0%	1.2%	4.9%	15.1%
West	20,000	yes	13	6.2%	1.4%	5.1%	16.3%
West	0	no	13	6.9%	1.5%	5.7%	18.1%
West	0	yes	13	6.9%	1.7%	5.9%	18.8%
West	20,000	no	18	3.8%	1.2%	4.5%	14.8%
West	20,000	yes	18	4.9%	1.6%	5.5%	20.6%
West	0	no	18	4.3%	1.4%	5.3%	17.9%
West	0	yes	18	5.1%	1.8%	6.3%	22.5%
East	20,000	no	9	4.3%	0.8%	2.9%	9.7%
East	20,000	yes	9	4.6%	0.9%	3.1%	10.5%
East	0	no	9	4.6%	0.9%	3.5%	10.4%
East	0	yes	9	4.5%	1.0%	3.6%	10.4%
East	20,000	no	13	4.1%	1.1%	3.7%	11.7%
East	20,000	yes	13	4.3%	1.1%	3.9%	12.9%
East	0	no	13	4.4%	1.2%	4.2%	13.0%
East	0	yes	13	4.4%	1.2%	4.4%	13.4%
East	20,000	no	18	3.3%	1.3%	4.8%	13.4%
East	20,000	yes	18	3.6%	1.5%	5.1%	15.7%
East	0	no	18	4.2%	1.6%	5.7%	17.7%
East	0	yes	18	4.3%	1.8%	5.9%	18.8%

Note: The simulated losses of expected lifetime consumption at age 40 are presented as reductions in net present values for scenarios 2-5 relative to scenario 1 (no health shock) that serves as reference scenario. The losses are simulated by endowments at age 40.

owner-occupied housing are likely to dissave less). Furthermore, a precautionary savings motive explains why individuals may save more than would be necessary for smoothing expected consumption over the life-cycle. Figure 4.2 shows that the realization of a persistent health shock may lead to substantial losses in expected lifetime consumption and that these losses depend substantially on the age when the health shock occurs. Net wealth and an employment history without any gaps appear to have protective effects on expected consumption. A persistent health shock at age 60 seems to induce comparatively small losses.

Table 4.3 shows simulated losses in NPVs of expected lifetime consumption at age 40 for scenarios 2-5 relative to scenario 1 (no health shock) that serves as reference scenario.[73] The difference in NPVs between scenario 2 (stochastic

[73]The NPVs of expected lifetime consumption at age 40 are presented in table 4.6 in the appendix.

health) and scenario 1 (no health shock) captures the magnitude of the health-related consumption risks that are uninsured by the German social security system. The differences between scenarios 3-5 and scenario 1 indicate the expected losses that are due to a persistent health shock at the ages 60, 55, and 45. The simulations suggest that expected health-related losses in lifetime consumption depend substantially on endowments at age 40 and range between 3% and 7%. The expected losses are larger for individuals without any net wealth at age 40 or with a lower level of education and are smaller in East than in West Germany. Individuals who lack net wealth may have delayed retirement in the absence of a health shock. Lower education goes along with a higher probability of bad health status and lower opportunity costs of early retirement.

The realization of a health risk leads to severe losses in expected lifetime consumption when the health shock occurs at an early stage of the life-cycle. For example, an individual in West Germany with a medium level of education (13 years), no gap in the employment history, and no net wealth at age 40 who experiences a persistent health shock at age 45 faces a loss in the NPV of expected lifetime consumption of 18.1%. The results of the simulations suggest that the German social security system may not sufficiently insure individuals against their health-related consumption risks. Of course, individuals can buy additional private disability insurance in order to achieve the desired level of insurance. These simulations merely show the limitations of the public social security system.

Old age poverty

This subsection focuses on distributional outcomes and examines the health-related risk of old age poverty. I take the EU definition of relative poverty as a reference that defines an individual as being at the "risk of poverty" when receiving a net income below 60% of the median net equivalent income. This indicates a threshold value of 816 EURO of monthly net income in 2005 (Statistisches Bundesamt (2008)).[74] In the following analysis, I define individuals as poor who experience a level of monthly consumption below 816 EURO. This is somewhat more conservative than the EU definition because monthly consumption after retirement may be higher than monthly income (as individuals consume out of their accumulated wealth).

[74] All nominal variables in my sample are adjusted to the purchasing power in 2005.

Table 4.4: Simulated risk of old age poverty at age 40

	Endowments at age 40			**Risk of old-age poverty**				
	Net wealth	5-year gap in empl.	Years of education	No health shock	Stochastic health	Shock at age 60	Shock at age 55	Shock at age 45
West	20,000	no	9	39.0%	45.5%	42.3%	45.9%	49.1%
West	20,000	yes	9	45.9%	53.1%	48.5%	52.1%	60.0%
West	0	no	9	42.7%	51.6%	46.6%	51.8%	59.2%
West	0	yes	9	49.8%	58.2%	52.8%	57.6%	67.4%
West	20,000	no	13	4.9%	8.4%	5.7%	7.6%	13.2%
West	20,000	yes	13	11.4%	17.7%	12.9%	15.9%	26.5%
West	0	no	13	5.4%	12.0%	6.3%	8.9%	24.7%
West	0	yes	13	13.8%	23.4%	15.6%	20.9%	40.8%
West	20,000	no	18	3.3%	4.4%	3.3%	4.5%	7.7%
West	20,000	yes	18	4.9%	7.8%	5.7%	7.4%	18.1%
West	0	no	18	3.5%	5.2%	3.5%	5.0%	12.6%
West	0	yes	18	5.3%	9.2%	6.0%	8.5%	26.9%
East	20,000	no	9	63.1%	66.4%	63.9%	65.0%	71.2%
East	20,000	yes	9	64.3%	68.4%	65.1%	66.3%	76.2%
East	0	no	9	63.3%	67.6%	64.0%	65.7%	77.1%
East	0	yes	9	65.0%	70.5%	65.7%	68.1%	82.4%
East	20,000	no	13	49.9%	54.1%	51.6%	54.7%	60.9%
East	20,000	yes	13	54.8%	59.3%	56.5%	59.3%	69.9%
East	0	no	13	52.8%	58.3%	55.0%	59.0%	67.0%
East	0	yes	13	57.4%	62.6%	59.1%	63.1%	77.2%
East	20,000	no	18	6.1%	8.3%	6.8%	8.9%	15.0%
East	20,000	yes	18	13.6%	16.8%	15.3%	18.4%	33.2%
East	0	no	18	6.4%	11.1%	7.3%	10.6%	28.9%
East	0	yes	18	16.6%	22.7%	19.0%	25.1%	50.5%

Note: The simulated risk of old age poverty at age 40 is presented for scenarios 1-5. The differences between scenarios 2-5 and scenario 1 (no health shock) indicate health-related changes in the risk of old-age poverty. The risks are simulated by endowments at age 40.

Table 4.4 presents the simulated risk of old age poverty by endowments at age 40. The differences between scenarios 2-5 and scenario 1 (no health shock) indicate health-related changes in the risk of old age poverty. For example, an individual with a medium level of education (13 years), no gap in the employment history, and no net wealth at age 40 who resides in West Germany faces a risk of old age poverty of 5.4% in the absence of health risks (scenario 1), while the risk rises to 12.0% under stochastic health (scenario 2), and amounts to 24.7% when the inividual experiences a persistent health shock at age 45 (scenario 5). In East Germany, an individual with the same endowments faces risks of 52.8%, 58.3%, and 67.0% in the respective scenarios. The magnitude of the risks and health-related changes in these risks depend substantially on an individual's endowments at age 40. The risk is greater in East than in West Germany and for individuals who lack net wealth or have experienced periods of non-employment. The health-related changes in the risk are sizeable for the endowments under consideration. Thus, there is a concern of health-related old age poverty that is uninsured by the German social security system.

4.7.2 Means-tested minimum pension benefits

Given the discussed concern of health-related old age poverty, policy makers might consider a public intervention. This subsection investigates a counterfactual reform of the statutory pension insurance scheme that protects individuals from the risk of old age poverty by introducing a minimum level of pension benefits at the - above defined - poverty line. However, such a reform raises a concern regarding an increase in abuse of the early retirement option and a decline in average pension age that is due to an increase in the attractivity of early retirement - in particular - for individuals with otherwise very low pension claims.[75] In the model, cheating is taken into account in the sense that individuals who are in bad health status may opt more often for early retirement even though some of these individuals are not work incapacitated such that their labor market participation decision is not restricted. For this reason, I resort to an idea of Golosov and Tsyvinski (2006) who argue that disability benefits should be means-tested in order to make it more unattractive to falsely claim benefits. The rationale behind this idea is that individuals need savings to smooth their consumption, but the more they save the more they are penalized

[75] There is also a countervailing effect because the induced reduction in the risk of future pension claims that depend on the labor market outcomes also leads to an increase in the option value of remaining in the labor force.

Table 4.5: Simulated changes in expected retirement age and expected lifetime consumption at age 40 through reform

	Endowments at age 40			**Δ E(retirement age)**		**Δ NPV**	
	Net wealth	5-year gap in employment	Years of education	means test	no means test	means test	no means test
West	20,000	no	9	-0.14	-0.93	0.31%	1.92%
West	20,000	yes	9	-0.15	-1.83	1.65%	3.19%
West	0	no	9	-0.13	-0.78	0.96%	2.67%
West	0	yes	9	-0.25	-1.51	2.11%	3.63%
West	20,000	no	13	-0.24	-0.60	-1.13%	-2.30%
West	20,000	yes	13	-0.22	-1.12	-0.97%	-2.00%
West	0	no	13	-0.26	-0.52	-1.23%	-2.08%
West	0	yes	13	-0.17	-0.82	-0.83%	-1.40%
West	20,000	no	18	0.08	0.12	0.27%	0.27%
West	20,000	yes	18	0.02	-0.17	0.33%	0.78%
West	0	no	18	0.05	0.07	0.30%	0.32%
West	0	yes	18	-0.03	-0.20	0.28%	0.46%
East	20,000	no	9	-0.27	-2.45	3.60%	6.16%
East	20,000	yes	9	-0.37	-3.68	3.83%	7.99%
East	0	no	9	-0.39	-1.80	3.18%	5.09%
East	0	yes	9	-0.41	-2.49	3.94%	6.82%
East	20,000	no	13	-0.22	-1.44	2.37%	4.17%
East	20,000	yes	13	-0.28	-2.49	2.73%	4.87%
East	0	no	13	-0.29	-1.23	2.71%	4.40%
East	0	yes	13	-0.33	-1.88	2.96%	4.75%
East	20,000	no	18	-0.30	-0.94	-2.46%	-3.66%
East	20,000	yes	18	-0.20	-1.33	-1.29%	-1.71%
East	0	no	18	-0.33	-0.75	-2.49%	-3.53%
East	0	yes	18	-0.21	-0.99	-1.03%	-1.19%

Note: The changes in expected retirement age and net present values of expected lifetime consumption at age 40 that are induced through the introduction of minimum pension benefits are simulated under stochastic health (scenario 2) both with and without a means test. The simulations are performed by endowments at age 40.

by the means test. This may prevent false claims of disability benefits if these benefits are not too generous relative to the wages that individuals can earn on the labor market. In the context of a pension scheme where the option of early retirement constitutes an insurance against work incapacity, this idea can be applied when introducing minimum pension benefits.

A means test may reduce the potential increase in abuse that is due to the introduction of minimum pension benefits and ensure that only individuals who are in need benefit from the reform. The scheme is set up as follows. If net pension benefits are below the minimum level (816 EURO of monthly consumption after retirement), they are increased to the minimum level, but the increase is made means-tested (as it is also the case for social assistance benefits). I use the model to simulate changes in expected retirement age by differing endowments at age 40. This is done analogously to the simulations in the previous subsection and under the assumption that the reform can be implemented without any increase in taxes or social security contributions (no budget neutrality). Given the level of means-tested social assistance benefits that can be claimed anyway (see section 3), any rise in taxes or social security contributions to finance the reform is small and, hence, behavioral responses to this rise would be small as well. The approach avoids the problem of choosing arbitrarily a financing scheme, where potential behavioral responses may depend on this choice.

Table 4.5 presents the simulated changes in expected retirement age by endowments at age 40 that are induced through the reform under stochastic health (scenario 2). For comparison, I also simulate a scheme of minimum pension benefits without a means test. The results indicate only a small decrease in expected retirement age (between 0 and 0.4 years depending on endowments) if the minimum pension benefits are means-tested. The lower the net wealth at age 40 and the lower the level of education the larger is the decrease. Even for individuals with a low level of education and no net wealth at age 40 the means test is highly relevant for expected retirement age. Without the means test, the decrease is substantial (between 0 and 3.7 years). Presumably, the difference between the behavioral responses for the two schemes would be even more pronounced when simulating the reform under budget neutrality. Overall, the introduction of means-tested minimum pension benefits appears to be a practicable approach in order to protect individuals from the risk of old age poverty.

4.8 Summary

This paper proposes a rich life-cycle model to investigate the health-related risks of consumption and old age poverty. The study demonstrates the good performance of an extension of the EM algorithm by estimating a complex dynamic model of health risks, labor market participation, early retirement, and wealth accumulation. The extension of the EM algorithm is used to obtain good starting values for a subsequent FIML estimation. I rely on the framework of a DPDC model that is estimated using data from the German Socio-Economic Panel Study. My analysis focuses on single males and, hence, allows abstracting from adjustments of the partner's labor supply and retirement behavior. The analyzed mechanisms are also relevant for couples and single females. Presumably, the health-related risks are smaller for couple households because adjustments of the partner's behavior may mitigate the economic consequences of a health shock. It follows that the health-related risks of single males suggest an upper bound for the respective risks of couple households.

I make life-cycle predictions for individuals by differing endowments at age 40. The model is used to simulate scenarios where health shocks do or do not occur at different ages during the life-cycle. A comparison of consumption paths and net present values of expected lifetime consumption between the scenarios sheds light on health-related consumption and poverty risks that are uninsured by the German social security system. The simulations suggest that expected health-related losses in lifetime consumption depend substantially on endowments and range between 3% and 7%. The expected losses are larger for individuals without any net wealth at age 40 or with a lower level of education and are smaller in East than in West Germany. The realization of a health risk leads to severe losses in expected lifetime consumption when the health shock occurs at an early stage of the life-cycle. The results suggest that the German social security system may not sufficiently insure individuals against their health-related consumption risks. Of course, individuals can buy additional private disability insurance in order to achieve the desired level of insurance. These simulations merely show the limitations of the public social security system.

I also use the life-cycle simulations to examine the health-related risk of old age poverty. The magnitude of the risk and health-related changes in this risk depend substantially on an individual's endowments. The risk is greater in

East Germany than in West Germany and for individuals who lack net wealth or have experienced periods of non-employment. The health-related changes in the risk are sizeable for the endowments under consideration. Thus, there is a concern of health-related old age poverty that is uninsured by the German social security system. Given this concern, I investigate a counterfactual reform of the German retirement scheme that protects individuals from the risk of old age poverty by introducing a minimum level of pension benefits at the poverty line. Since minimum pension benefits may go along with an increase in abuse of the early retirement option and a decrease in average pension age, I examine whether a means test may reduce the potential increase in abuse. The simulations indicate only a small decrease in expected retirement age (between 0 and 0.4 years depending on endowments) if the minimum pension benefits are means-tested. Without the means test, the decrease is substantial (between 0 and 3.7 years). Overall, the introduction of means-tested minimum pension benefits appears to be a practicable approach in order to protect individuals from the risk of old age poverty.

4.9 Appendix

Table 4.6 shows the net present values of expected lifetime consumption by endowments at age 40 for scenarios 1-5.

Table 4.6: NPVs of expected lifetime consumption at age 40

	Endowments at age 40			**NPVs (EURO)**				
	Net wealth	5-year gap in empl.	Years of education	No health shock	Stochastic health	Shock at age 60	Shock at age 55	Shock at age 45
West	20,000	no	9	353,420	332,410	349,510	339,610	310,960
West	20,000	yes	9	338,440	318,060	334,540	324,960	294,380
West	0	no	9	338,670	317,440	334,490	323,770	293,060
West	0	yes	9	324,860	304,530	320,760	310,260	279,500
West	20,000	no	13	482,280	453,550	476,490	458,730	409,380
West	20,000	yes	13	451,030	423,160	444,590	428,170	377,700
West	0	no	13	462,130	430,130	455,000	435,720	378,470
West	0	yes	13	428,440	398,960	421,260	403,190	347,730
West	20,000	no	18	578,540	556,580	571,530	552,320	492,900
West	20,000	yes	18	563,640	535,910	554,750	532,520	447,480
West	0	no	18	556,500	532,540	548,650	527,270	457,150
West	0	yes	18	543,100	515,170	533,330	509,020	420,670
East	20,000	no	9	283,880	271,580	281,670	275,550	256,370
East	20,000	yes	9	276,520	263,780	274,070	268,020	247,510
East	0	no	9	276,120	263,500	273,620	266,390	247,300
East	0	yes	9	268,300	256,280	265,600	258,780	240,460
East	20,000	no	13	326,060	312,570	322,590	314,060	287,880
East	20,000	yes	13	314,520	301,020	310,970	302,200	273,850
East	0	no	13	313,830	300,180	310,060	300,640	272,940
East	0	yes	13	303,120	289,930	299,380	289,800	262,410
East	20,000	no	18	470,520	455,080	464,400	448,100	407,470
East	20,000	yes	18	437,580	421,650	430,910	415,300	368,980
East	0	no	18	450,310	431,200	443,080	424,800	370,410
East	0	yes	18	415,010	397,230	407,660	390,440	336,900

Note: Expected lifetime consumption at age 40 is presented in terms of net present values for scenarios 1-5. The simulations are performed by endowments at age 40.

List of Tables

List of Figures

Bibliography

Adda, J. and R. W. Cooper (2003): *Dynamic Economics: Quantitative Methods and Applications*, The MIT Press.

Adda, J., C. Dustmann, C. Meghir, and J.-M. Robin (2009): "Career progression and formal versus on-the-job training," IFS Working Papers W09/06, Institute for Fiscal Studies.

Aguirregabiria, V. and P. Mira (2010): "Dynamic discrete choice structural models: A survey," *Journal of Econometrics*, 156, 38–67.

Akay, A. (2009): "Dynamics of the Employment Assimilation of First-Generation Immigrant Men in Sweden: Comparing Dynamic and Static Assimilation Models with Longitudinal Data," IZA Discussion Paper 4655.

——— (2011): "Finite-sample comparison of alternative methods for estimating dynamic panel data models," *Journal of Applied Econometrics*.

Albert, J. H. and S. Chib (1993): "Bayesian Analysis of Binary and Polychotomous Response Data," *Journal of the American Statistical Association*, 88, 669–679.

Albouy, V. and L. Lequien (2009): "Does compulsory education lower mortality?" *Journal of Health Economics*, 28, 155–168.

Albrecht, M., S. Loos, and G. Schiffhorst (2007): "Sozioökonomische Situation von Personen mit Erwerbsminderung: eine Analyse auf Basis von SOEP-Daten," in *Deutsche Rentenversicherung*, 10, 621–640.

Allenby, G. (1997): "An introduction to hierarchical Bayesian modelling," *Tutorial Notes*.

Allenby, G. M. and P. J. Lenk (1994): "Modeling Household Purchase Behavior with Logistic Normal Regression," *Journal of the American Statistical Association*, 89, 1218–1231.

Andersen, H. H., A. Mühlbacher, M. Nübling, J. Schupp, and G. G. Wagner (2007): "Computation of standard values for physical and mental health scale scores using the SOEP version of SF-12v2," *Schmollers Jahrbuch*, 127, 171–182.

Angrist, J. and G. Imbens (1995): "Two-stage least squares estimation of average causal effects in models with variable treatment intensity," *Journal of the American Statistical Association*, 90, 431–442.

Arcidiacono, P. (2004): "Ability sorting and the returns to college major," *Journal of Econometrics*, 121, 343–375.

Arcidiacono, P. and J. B. Jones (2003): "Finite Mixture Distributions, Sequential Likelihood and the EM Algorithm," *Econometrica*, 71, 933–946.

Arcidiacono, P., H. Sieg, and F. Sloan (2007): "Living Rationally Under The Volcano? An Empirical Analysis Of Heavy Drinking And Smoking," *International Economic Review*, 48, 37–65.

Becker, G. S., J. S. Duesenberry, and B. Okun (1960): "An economic analysis of fertility," in *Demographic and economic change in developed countries*, ed. by National Bureau of Economic Research, Princeton: Princeton University Press, 225–256.

Becker, G. S. and H. G. Lewis (1973): "On the interaction between the quantity and quality of children," *Journal of Political Economy*, 81, 279–288.

Becker, G. S. and C. B. Mulligan (1997): "The Endogenous Determination of Time Preference," *The Quarterly Journal of Economics*, 112, 729–758.

Benitez-Silva, H. and H. Ni (2008): "Health status and health dynamics in an empirical model of expected longevity," *Journal of Health Economics*, 27, 564–584.

Black, S. E. and P. J. Devereux (2011): "Recent developments in intergenerational mobility," in *Handbook of Labor Economics 4B*, ed. by O. Ashenfelter and D. Card, Elsevier, chap. 16, 1487–1541.

Black, S. E., P. J. Devereux, and K. G. Salvanes (2005): "The more the merrier? The effect of family size and birth order on children's education," *Quarterly Journal of Economics*, 120, 669–700.

Blau, D. M. and D. B. Gilleskie (2008): "The Role Of Retiree Health Insurance In The Employment Behavior Of Older Men," *International Economic Review*, 49, 475–514.

Boersch-Supan, A. (2001): "Incentive Effects of Social Security Under an Uncertain Disability Option," in *Themes in the Economics of Aging*, ed. by D. Wise, 281–310.

Bound, J., M. Schoenbaum, T. R. Stinebrickner, and T. Waidmann (1999): "The dynamic effects of health on the labor force transitions of older workers," *Labour Economics*, 6, 179–202.

Boyles, R. A. (1983): "On the Convergence of the EM Algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, 45, 47–50.

Card, D. (1999): "The causal effect of education on earnings," in *Handbook of Labor Economics 3A*, ed. by O. C. Ashenfelter and D. Card, New York: Elsevier, vol. 3, 1801–1863.

Carneiro, P., C. Crawford, and A. Goodman (2007): "The Impact of Early Cognitive and Non-Cognitive Skills on Later Outcomes," CEE Discussion Papers 0092.

Carneiro, P., C. Meghir, and M. Parey (2013): "Maternal Education, Home Environments, and the Development of Children and Adolescents," *Journal of the European Economic Association*, 11, 123–160.

Case, A., D. Lubotsky, and C. Paxson (2002): "Economic Status and Health in Childhood: The Origins of the Gradient," *American Economic Review*, 92, 1308–1334.

Case, A. and C. Paxson (2005): "Sex Differences in Morbidity and Mortality," *Demography*, 42, 189–214.

Chevalier, A. and V. O'Sullivan (2007): "Mother's education and birth weight," IZA Discussion Paper 2640.

Chiappori, P. (1988): "Rational Household Labor Supply," *Econometrica*, 56, 63–89.

Clark, D. and H. Royer (2010): "The Effect of Education on Adult Health and Mortality: Evidence from Britain," NBER Working Papers 16013.

CONEUS, K. AND C. K. SPIESS (2012): "The intergenerational transmission of health in early childhood: Evidence from the German Socio-Economic Panel study." *Economics and Human Biology*, 10, 89–97.

CURRIE, J. (2009): "Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development," *Journal of Economic Literature*, 47, 87–122.

CURRIE, J. AND E. MORETTI (2003): "Mother's education and the intergenerational transmission of human capital: Evidence from college openings," *Quarterly Journal of Economics*, 118, 1495–1532.

CUTLER, D. M. AND A. LLERAS-MUNEY (2008): *Education and Health: Evaluating theories and evidence*, Russell Sage Press.

——— (2010): "Understanding differences in health behaviors by education," *Journal of Health Economics*, 29, 1–28.

DARON ACEMOGLU, J. A. (2000): *How Large are Human-Capital Externalities? Evidence from Compulsory-Schooling Laws*, MIT Press, NBER Macroeconomics Annual, Volume 15.

DAVID, M. C. AND A. LLERAS-MUNEY (2012): "Education and Health: Insights from International Comparisons," NBER Working Papers 17738.

DE NARDI, M., E. FRENCH, AND J. B. JONES (2010): "Why Do the Elderly Save? The Role of Medical Expenses," *Journal of Political Economy*, 118, 39–75.

DE WALQUE, D. (2007): "Does education affect smoking behaviors?: Evidence using the Vietnam draft as an instrument for college education," *Journal of Health Economics*, 26, 877–895.

DEATON, A. (2002): "Policy Implications of the Gradient of Health and Wealth," *Health Affairs*, 21, 13–30.

——— (2003): "Health, Inequality, and Economic Development," *Journal of Economic Literature*, 41, 113–158.

DEUTSCHE RENTENVERSICHERUNG (2011): *Rentenversicherung in Zeitreihen*, vol. 22 of *DRV-Schriften*.

Deutsches Krebsforschungszentrum (2004): *Rauchen und soziale Ungleichheit - Konsequenzen für die Tabakkontrollpolitik [Smoking and Health Inequality - Implications for tobacco control policy]*, Heidelberg.

Disney, R., C. Emmerson, and M. Wakefield (2006): "Ill health and retirement in Britain: A panel data-based analysis," *Journal of Health Economics*, 25, 621–649.

Dube, J.-P., G. J. Hitsch, and P. E. Rossi (2010): "State dependence and alternative explanations for consumer inertia," *Rand Journal of Economics*, 41, 417–445.

Dustmann, C. (2004): "Parental background, secondary school track choice, and wages," *Oxford Economic Papers*, 56, 209–230.

Eckstein, Z. and K. I. Wolpin (1989): "Dynamic Labour Force Participation of Married Women and Endogenous Work Experience," *Review of Economic Studies*, 56, 375–90.

Fan, J. and I. Gijbels (1996): *Local polynomial modelling and its applications*, London: Chapman and Hall.

Fitzenberger, B., A. Osikominu, and M. Paul (2010): "The Heterogeneous Effects of Training Incidence and Duration on Labor Market Transitions," IZA Discussion Paper 5269.

Fletcher, J. M. and D. E. Frisvold (2012): "The long run health returns to college quality," *Review of Economics of the Household.*

French, E. (2005): "The Effects of Health, Wealth, and Wages on Labour Supply and Retirement Behaviour," *Review of Economic Studies*, 72, 395–427.

French, E. and J. B. Jones (2011): "The Effects of Health Insurance and Self-Insurance on Retirement Behavior," *Econometrica*, 79, 693–732.

Fuchs, V. (1982): "Time Preference and Health: An Explanatory Study," in *Economic aspects of health*, ed. by V. Fuchs, University of Chicago Press, 93–120.

Fuchs-Schuendeln, N. (2008): "The Response of Household Saving to the Large Shock of German Reunification," *American Economic Review*, 98, 1798–1828.

Fuchs-Schuendeln, N. and M. Schuendeln (2005): "Precautionary Savings and Self-Selection: Evidence from the German Reunification "Experiment"," *The Quarterly Journal of Economics*, 120, 1085–1120.

Ganzeboom, H., P. De Graaf, and D. Treiman (1992): "A standard international socio-economic index of occupational status," *Social Science Research*, 21, 1–56.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (1995): *Bayesian Data Analysis*, Chapman and Hall.

Golosov, M. and A. Tsyvinski (2006): "Designing Optimal Disability Insurance: A Case for Asset Testing," *Journal of Political Economy*, 114, 257–279.

Gourinchas, P.-O. and J. A. Parker (2002): "Consumption Over the Life Cycle," *Econometrica*, 70, 47–89.

Grossman, M. (1972): "On the Concept of Health Capital and the Demand for Health," *Journal of Political Economy*, 80, 223–55.

——— (2006): "Education and Nonmarket Outcomes," in *Handbook of the Economics of Education*, ed. by E. Hanushek and F. Welch, Elsevier, vol. 1, chap. 10, 577–633.

Gugl, E. and L. Welling (2011): "Time with sons and daughters," *Review of Economics of the Household*, 10, 277–298.

Haan, P. (2010): "A Multi-state model of state dependence in labor supply: Intertemporal labor supply effects of a shift from joint to individual taxation," *Labour Economics*, 17, 323–335.

Haan, P., D. Kemptner, and A. Uhlendorff (2012): "Bayesian Procedures as a Numerical Tool for the Estimation of Dynamic Discrete Choice Models," DIW Discussion Papers 1210.

Haan, P. and M. Myck (2009): "Dynamics of health and labor market risks," *Journal of Health Economics*, 28, 1116–1125.

Haan, P. and V. Prowse (2011): "Longevity, Life-cycle Behavior and Pension Reform," DIW Discussion Papers 1140.

Haan, P. and A. Uhlendorff (2012): "Intertemporal Labor Supply and Involuntary Unemployment," *Empirical Economics*, forthcoming.

Hastings, W. K. (1970): "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, 57, 97–109.

Heckman, J. (1981a): "Heterogeneity and State Dependence," in *Studies in Labor Markets*, ed. by S. Rosen, Chicago, IL: Chicago Press, 91–139.

——— (1981b): "The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process," in *Structural Analysis of Discrete Data with Econometric Applications*, ed. by C. Manski and D. McFadden, MIT Press, chap. 4, 179–195.

——— (1981c): "Statistical Models for Discrete Panel Data," in *Structural Analysis of Discrete Data with Econometric Applications*, ed. by C. Manski and D. McFadden, MIT Press, 114–178.

Heckman, J. and B. Singer (1984): "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, 52, 271–320.

Himmelreicher, R. K. and A. Stuchlik (2008): "Entwicklung und Verteilung von Entgeltpunkten in der gesetzlichen Rentenversicherung," in *Deutsche Rentenversicherung*, 6, 532–547.

Imai, S., N. Jain, and A. Ching (2009): "Bayesian Estimation of Dynamic Discrete Choice Models," *Econometrica*, 77, 1865–1899.

Jepsen, L. K. (2005): "The relationship between wife's education and husband's earnings: Evidence from 1960 to 2000," *Review of Economics of the Household*, 3, 197–214.

Jürges, H., S. Reinhold, and M. Salm (2011): "Does schooling affect health behavior? Evidence from the educational expansion in Western Germany," *Economics of Education Review*, 30, 862–872.

Jürges, H. and K. Schneider (2007): "What Can Go Wrong Will Go Wrong: Birthday Effects and Early Tracking in the German School System," CESifo Working Paper Series 2055.

Keane, M. P. and K. I. Wolpin (1994): "The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence," *The Review of Economics and Statistics*, 76, 648–72.

——— (2001): "Estimating Welfare Effects Consistent with Forward-Looking Behavior: Part II: Empirical Results," *Journal of Human Resources*, 37, 600–622.

Kemptner, D., H. Juerges, and S. Reinhold (2011): "Changes in compulsory schooling and the causal effect of education on health: Evidence from Germany," *Journal of Health Economics*, 30, 340–354.

Kemptner, D. and J. Marcus (2013): "Spillover effects of maternal education on child's health and health behavior," *Review of Economics of the Household*, 11, 29–52.

Kenkel, D. S. (1991): "Health Behavior, Health Knowledge, and Schooling," *Journal of Political Economy*, 99, 287–305.

Lamerz, A., J. Kuepper-Nybelen, C. Wehle, N. Bruning, G. Trost-Brinkhues, H. Brenner, J. Hebebrand, and B. Herpertz-Dahlmann (2005): "Social class, parental education, and obesity prevalence in a study of six-year-old children in Germany," *International Journal of Obesity*, 29, 373–380.

Lancaster, T. (1990): *The Econometric Analysis of Transition Data*, Cambridge: Cambridge University Press.

Lindeboom, M., A. Llena-Nozal, and B. van der Klaauw (2009): "Parental education and child health: Evidence from a schooling reform." *Journal of Health Economics*, 28, 109–131.

Little, R. J. and D. B. Rubin (2002): *Statistical analysis with missing data*, Hoboken, NJ: Wiley, 2nd ed.

Lleras-Muney, A. (2005): "The Relationship Between Education and Adult Mortality in the United States," *Review of Economic Studies*, 72, 189–221.

Loureiro, M. L., A. Sanz-de Galdeano, and D. Vuri (2010): "Smoking habits: Like father, like son, like mother, like daughter?" *Oxford Bulletin of Economics and Statistics*, 72, 717–743.

Low, H. and L. Pistaferri (2011): "Disability Risk, Disability Insurance and Life Cycle Behavior," Working paper.

Martius, J., T. Steck, M. Oehler, and K. Wulf (1998): "Risk factors associated with preterm ($< 37+$ 0 weeks) and early preterm birth ($< 32+$ 0 weeks): Univariate and multivariate analysis of 106 345 singleton births from the 1994 statewide perinatal survey of Bavaria," *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 80, 183–189.

McCrary, J. and H. Royer (2011): "The effect of female education on fertility and infant health: Evidence from school entry policies using exact date of birth," *American Economic Review*, 101, 158–195.

McCulloch, R. and P. E. Rossi (1994): "An exact likelihood analysis of the multinomial probit model," *Journal of Econometrics*, 64, 207–240.

McFadden, D. and K. Train (2000): "Mixed MNL models for discrete response," *Journal of Applied Econometrics*, 15, 447–470.

McIntire, D., S. Bloom, B. Casey, and K. Leveno (1999): "Birth weight in relation to morbidity and mortality among newborn infants," *New England Journal of Medicine*, 340, 1234–1238.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953): "Equation of State Calculations by Fast Computing Machines," *Journal of Chemical Physics*, 21, 1087–1092.

Mokdad, A., J. Marks, D. Stroup, and J. Gerberding (2004): "Actual causes of death in the united states, 2000," *The Journal of the American Medical Association*, 291, 1238–1245.

Naeye, R. (1964): "The fetal and neonatal development of twins," *Pediatrics*, 33, 546–553.

Nayga, R. (2000): "Schooling, health knowledge and obesity," *Applied Economics*, 32, 815–822.

Oreopoulos, P. (2006): "Estimating Average and Local Average Treatment Effects of Education when Compulsory Schooling Laws Really Matter," *American Economic Review*, 96, 152–175.

Oreopoulos, P., M. E. Page, and A. H. Stevens (2008): "The intergenerational effects of worker displacement," *Journal of Labor Economics*, 26, 455–483.

Piopiunik, M. (2011): "Intergenerational transmission of education and mediating channels: Evidence from compulsory schooling reforms in Germany," *Ifo Working Paper*, 107.

Pischke, J.-S. (2007): "The Impact of Length of the School Year on Student Performance and Earnings: Evidence From the German Short School Years," *Economic Journal*, 117, 1216–1242.

Pischke, J.-S. and T. von Wachter (2005): "Zero Returns to Compulsory Schooling In Germany: Evidence and Interpretation," NBER Working Papers 11414.

——— (2008): "Zero Returns to Compulsory Schooling in Germany: Evidence and Interpretation," *The Review of Economics and Statistics*, 90, 592–598.

Powdthavee, N. (2010): "Does Education Reduce the Risk of Hypertension? Estimating the Biomarker Effect of Compulsory Schooling in England," IZA Discussion Papers 4847.

Prowse, V. (2010): "Modeling Employment Dynamics with State Dependence and Unobserved Heterogeneity," IZA Discussion Paper 4489.

Regier, D. A., M. Ryan, E. Phimister, and C. A. Marra (2009): "Bayesian and classical estimation of mixed logit: An application to genetic testing," *Journal of Health Economics*, 28.

Reinhold, S. and H. Jürges (2010): "Secondary school fees and the causal effect of schooling on health behavior," *Health Economics*, 19, 994–1001.

Riphahn, R. (2008): "The effect of secondary school fees on educational attainment," LASER Discussion Papers 21.

Rosenzweig, M. and T. Schultz (1981): *Education and Household Production of Child Health*, Russell Sage Press, 382–387.

Rubin, D. B. (1976): "Inference and missing data," *Biometrika*, 63, 581–592.

Rust, J. (1987): "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 55, 999–1033.

RUST, J. AND C. PHELAN (1997): "How Social Security and Medicare Affect Retirement Behavior in a World of Incomplete Markets," *Econometrica*, 65, 781–832.

SANDER, W. (1995): "Schooling and quitting smoking," *Review of Economics and Statistics*, 77, 191–199.

SILLES, M. A. (2009): "The causal effect of education on health: Evidence from the United Kingdom," *Economics of Education Review*, 28, 122–128.

STARRELS, M. (1994): "Gender differences in parent-child relations," *Journal of Family Issues*, 15, 148.

STATACORP (2009): "lpoly," in *Stata: Release 11. Statistical Software.*, College Station, TX: StataCorp LP.

STATISTISCHES BUNDESAMT (2008): "Leben in Europa (EU-SILC)," in *Wirtschaftsrechnungen*, vol. 15.

STEINER, V., K. WROHLICH, J. GEYER, AND P. HAAN (2008): "Documentation of the Tax-Benefit Microsimulation Model STSM: Version 2008," *Data Documentation 31.*

TRAIN, K. (2001): "A Comparison of Hierarchical Bayes and Maximum Simulated Likelihood for Mixed Logit," *Technical Report.*

——— (2009): *Discrete Choice Models using Simulation*, Cambridge, UK: Cambridge University Press, second ed.

TROSKE, K. AND A. VOICU (2010): "Joint Estimation of Sequential Labor Force Participation and Fertility Decisions Using Markov Chain Monte Carlo Techniques," *Labour Economics*, 17, 150–169.

VAN DEN BERG, G. J. (2001): "Duration Models: Specification, Identification, and Multiple Durations," in *Handbook of Econometrics, Vol.5*, ed. by J. J. Heckman and E. Leamer, Amsterdam: North-Holland, 3381–3460.

VAN DER KLAAUW, W. AND K. I. WOLPIN (2008): "Social security and the retirement and savings behavior of low-income households," *Journal of Econometrics*, 145, 21–42.

VAN SOEST, A. (1995): "Structural Models of Family Labor Supply: A Discrete Choice Approach," *Journal of Human Resources*, 30, 63–88.

VERBRUGGE, L. (1989): "The twain meet: empirical explanations of sex differences in health and mortality," *Journal of Health and Social Behavior*, 30, 282–304.

WAGNER, G. G., J. R. FRICK, AND J. SCHUPP (2007): "The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements," *Journal of Applied Social Science Studies*, 127, 139–169.

WOLPIN, K. I. (1984): "An Estimable Dynamic Stochastic Model of Fertility and Child Mortality," *Journal of Political Economy*, 92, 852–74.

WOOLDRIDGE, J. M. (2002): *Econometric analysis of cross section and panel data*, The MIT press.

——— (2005): "Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity," *Journal of Applied Econometrics*, 20, 39–54.

WORLD HEALTH ORGANIZATION (2003): *Social determinants of health. The solid facts*, http://www.euro.who.int/document/e81384.pdf, accessed March 16, 2010.

WU, C. (1983): "On the Convergence Properties of the EM Algorithm," *The Annals of Statistics*, 11, 95–103.

ZEGER, S. L. AND M. R. KARIM (1991): "Generalized Linear Models With Random Effects; A Gibbs Sampling Approach," *Journal of the American Statistical Association*, 86, pp. 79–86.

ZELLNER, A. AND P. E. ROSSI (1984): "Bayesian analysis of dichotomous quantal response models," *Journal of Econometrics*, 25, 365–393.

Summary

Chapter 1: The first paper contributes to the growing literature on identifying the causal link between education and health and health-related behavior. Numerous studies have indeed documented a strong positive empirical association between education and health. Using data from several German Microcensuses, we exploit changes in years of compulsory schooling in West German federal states that took effect between 1949 and 1969 to estimate the causal effect of years in school on long-term illness, work disability, BMI (and overweight/obesity) and current and former smoking measured in 1989 to 2003. We find evidence for a strong and significant negative causal effect of years of schooling on long-term illness and work disability among men. Our IV estimates are slightly larger than OLS but in contrast to some of the existing literature they tend to be in the same range. For women, however, we do not find any significant causal effects on health status. We also find some evidence for negative causal effects of education on male body weight and somewhat less convincing support for a negative effect on female body weight.

We address some possible concerns about the validity of our results. The first robustness check - exploiting the fact that four large contiguous states have lengthened compulsory schooling in the same year - suggests that migration is unlikely to compromise our estimation results. Further, we address the concern that unobserved state-specific trends which could be correlated with the instrument are biasing our results. In our most flexible specification, we allow for (common) cohort fixed effects and cubic state-specific cohort trends. These robustness checks do not alter our main conclusions. Next, we address the concern that changes in mandatory schooling laws could also affect the likelihood that a student leaves school without a degree or induces track changes that actually reduce the number of years in school. Such behavior would violate the monotonicity assumption made in the presence of heterogeneous treatment effects. Regression results indicate very small effects on track choice that are unlikely to threaten the validity of our main results. Finally, we allow for the

fact that students affected by the mandatory schooling reform also underwent so-called short school years. Recoding our endogenous variable to account for short school years does not alter our results either.

A link between education and health outcomes via better health inputs in terms of less smoking is not supported by our data. However, overweight and obesity can be interpreted not only as indicators of future health problems but also as indicators of past health behavior. Thus, our findings on weight indirectly support the health input argument, although more so for men than for women. Changes in occupation from manual to non-manual could be an explanation. Further, we have not looked at income or wages in our paper. Considering existing evidence - using an identical identification strategy - changes in mandatory schooling had no causal effect on wages (Pischke and von Wachter (2008)), the link between education and health via higher income (and thus favorable living conditions) appears unlikely. Finally, it must also be noted that our parameters only identify the effect of education for compliers to the specific reforms of raising mandatory school leaving age. Interventions at other stages of the life-cycle or more specific interventions might have stronger and more systematic effects on health outcomes and health behavior.

Chapter 2: The second paper investigates the effects of maternal education on child's health and health behavior in Germany. A quantification of such intergenerational links is not only relevant regarding optimal investments into education, but also relates to social mobility. We add to the literature by applying an IV strategy that works for the sample size of common household panels, by considering a wide range of outcomes for both newborns and adolescents, and by investigating possible channels of the estimated effects. We instrument maternal education by the number of her siblings while conditioning on characteristics of her parents, the child's grandparents. For this purpose, we draw on a rich household survey, the German Socio-Economic Panel Study, containing detailed information about three generations. We argue that, given the grandparents' characteristics, the number of the mother's siblings generates variation in maternal years of education that is exogenous regarding her child's health and health behavior.

For newborns, we find a significant negative association between maternal education and the probability of preterm maturity. Although the effect on preterm maturity increases in size when estimating the IV-Probit model, the effect turns insignificant. It may be that the size of our newborns sample is not large enough to detect existing effects with the IV approach. We find no evidence for an effect of maternal education on low birth weight. For adolescents, the IV approach suggests strong and significant effects on health-related behavior for daughters. One more year of maternal education is estimated to reduce the daughter's probability of smoking at age 18/19 by 7.4 percentage points and to increase the daughter's likelihood of doing sports at least once a week by 7.5 percentage points. However, we do not obtain significant effects of maternal education on sons' health behavior. For both sexes, we do not find any effects on child's physical health or overweight. Concerning the relevance of the instrument, we find all respective first stage F-statistics to exceed the critical value of 10.

We demonstrate the robustness of our IV estimates by sequentially introducing grandparental characteristics. Also the results do not change, when we control for further variables (like mother's health, health behavior, and fertility) that are possibly consequences of maternal education. The results are robust to only considering mothers with one, two or three siblings as well as to functional form assumptions of the first stage or distributional assumptions of the error terms. Investigating possible channels of the estimated effects, our findings do not suggest that mother's health behavior, assortative mating, or household income explain the effects on daughter's health behavior. However, when including the child's schooling track as an additional control variable in our IV approach, the effect of maternal education on daughter's smoking behavior disappears. Thus, maternal education seems to affect child's health behavior by affecting schooling track.

Chapter 3: The third paper provides evidence for the advantage of using Bayesian estimation procedures instead of classical maximum likelihood estimation for the estimation of dynamic discrete choice models. These models usually require a general specification of unobserved preference heterogeneity and therefore often relatively complex estimation routines need to be applied. We estimate a dynamic discrete choice model of female labor supply with three distinct states and different specifications of unobserved heterogeneity. Our analysis is based on longitudinal data from the German Socioeconomic Panel.

The empirical comparison leads to two important conclusions which are highly relevant for practitioners estimating dynamic discrete choice models with unobserved heterogeneity. First, when considering a multivariate normal distribution for the unobserved heterogeneity both approaches, the MCMC estimator and the MSL estimation, yield almost identical results. This shows that for a finite sample of the size which is typical for common household panels, our findings are in line with the asymptotic results of the Bernstein-von Mises Theorem. Hence, the Bayesian estimates can be given a classical interpretation. The second finding demonstrates the advantage of using Bayesian estimation procedures. We show that when imposing distributional assumptions which are consistent with economic theory, e.g. log-normally distributed consumption preferences, the Bayesian method performs well and provides reasonable estimates, while the MSL estimator does not converge. These results indicate that Bayesian procedures can be a beneficial tool for the estimation of dynamic discrete choice models.

Chapter 4: The fourth paper proposes a rich life-cycle model to investigate the health-related risks of consumption and old age poverty. The study demonstrates the good performance of an extension of the EM algorithm by estimating a complex dynamic model of health risks, labor market participation, early retirement, and wealth accumulation. The extension of the EM algorithm is used to obtain good starting values for a subsequent FIML estimation. I rely on the framework of a DPDC model that is estimated using data from the German Socio-Economic Panel Study. My analysis focuses on single males and, hence, allows abstracting from adjustments of the partner's labor supply and retirement behavior. The analyzed mechanisms are also relevant for couples and single females. Presumably, the health-related risks are smaller for couple households because adjustments of the partner's behavior may mitigate the economic consequences of a health shock. It follows that the health-related risks of single males suggest an upper bound for the respective risks of couple households.

I make life-cycle predictions for individuals by differing endowments at age 40. The model is used to simulate scenarios where health shocks do or do not occur at different ages during the life-cycle. A comparison of consumption paths and net present values of expected lifetime consumption between the scenarios sheds light on health-related consumption and poverty risks that are uninsured by the German social security system. The simulations suggest that expected health-related losses in lifetime consumption depend substantially on endowments and range between 3% and 7%. The expected losses are larger for individuals without any net wealth at age 40 or with a lower level of education and are smaller in East than in West Germany. The realization of a health risk leads to severe losses in expected lifetime consumption when the health shock occurs at an early stage of the life-cycle. The results suggest that the German social security system may not sufficiently insure individuals against their health-related consumption risks. Of course, individuals can buy additional private disability insurance in order to achieve the desired level of insurance. These simulations merely show the limitations of the public social security system.

I also use the life-cycle simulations to examine the health-related risk of old age poverty. The magnitude of the risk and health-related changes in this risk depend substantially on an individual's endowments. The risk is greater in East Germany than in West Germany and for individuals who lack net wealth or have experienced periods of non-employment. The health-related changes in the risk are sizeable for the endowments under consideration. Thus, there is a concern of health-related old age poverty that is uninsured by the German social security system. Given this concern, I investigate a counterfactual reform of the German retirement scheme that protects individuals from the risk of old age poverty by introducing a minimum level of pension benefits at the poverty line. Since minimum pension benefits may go along with abuse of the early retirement option and a decrease in average pension age, I examine whether a means test may reduce the potential increase in abuse. The simulations indicate only a small decrease in expected retirement age (between 0 and 0.4 years depending on endowments) if the minimum pension benefits are means-tested. Without the means test, the decrease is substantial (between 0 and 3.7 years). Overall, the introduction of means-tested minimum pension benefits appears to be a practicable approach in order to protect individuals from the risk of old age poverty.

German Summary

Kapitel 1: Das erste Kapitel leistet einen Beitrag zur wachsenden Literatur, die sich um eine Identifikation des Kausalzusammenhangs zwischen Bildung und Gesundheit sowie gesundheitsbezogenem Verhalten bemüht. In der Tat haben zahlreiche Studien einen starken positiven empirischen Zusammenhang zwischen Bildung und Gesundheit dokumentiert. Unter Verwendung mehrer Erhebungsjahre des deutschen Mikrozensus nutzen wir Änderungen in den Pflichtschulregelungen der westdeutschen Bundesländer aus, die sich zwischen 1949 und 1969 ereignet haben, um den kausalen Effekt der Schuljahre auf die Wahrscheinlichkeit einer chronischen Erkrankung, einer Arbeitsunfähigkeit, den BMI (und Übergewicht und Fettleibigkeit) und das Raucherverhalten in den Jahren von 1989 bis 2003 zu schätzen. Wir finden Evidenz für einen starken und signifikanten negativen Effekt der Bildungsjahre auf die Wahrscheinlichkeit einer chronischen Erkrankung und einer Arbeitsunfähigkeit bei Männern. Unsere IV Schätzungen sind dabei etwas größer als die OLS Schätzungen, befinden sich aber im Gegensatz zu einigen anderen Studien in der gleichen Größenordnung. Für Frauen finden wir keinen signifikanten Effekt auf den Gesundheitsstatus. Außerdem finden wir Evidenz für einen negativen Effekt der Bildung auf das männliche Körpergewicht und etwas schwächere Evidenz für einen negativen Effekt auf das weibliche Körpergewicht.

Wir untersuchen einige mögliche Beeinträchtigungen der Validität unserer Ergebnisse. Die erste Robustheitsprüfung, welche die Tatsache ausnutzt, dass vier große aneinander angrenzende Bundesländer die Pflichtschulzeit im selben Jahr verlängert haben, weist darauf hin, dass eine Beeinträchtigung unserer Schätzergebnisse durch Migration unwahrscheinlich ist. Zudem adressieren wir das Problem, dass unbeobachtete bundesländerspezifische Trends, die mit dem Instrument korreliert sein könnten, zu einer Verzerrung der Ergebnisse führen könnten. In der flexibelsten Spezifikation, kontrollieren wir für (gemeinsame) fixe Kohorteneffekte und kubische (d.h. Polynom dritter Ordnung) länderspezifische Trends. Diese Robustheitsprüfungen ändern nichts an unseren Schlussfolgerungen. Des Weiteren untersuchen wir das Problem, dass eine Erhöhung der Pflichtschulzeit auch die Wahrscheinlichkeit erhöhen könnte, dass ein Schüler die Schule ohne Abschluss verlässt oder es zu Änderungen in der Wahl des Bildungspfads kommt, die effektiv zu einer Reduzierung der Schuljahre führen. Ein solches Verhalten würde die Monotonitätsannahme verletzen, die bei Vorliegen heterogener Effekte getroffen werden muss. Regressionsergebnisse zeigen sehr kleine Effekte auf die Wahl des Bildungspfads an, so dass eine Beeinträchtigung der Validität unserer Ergebnisse unwahrscheinlich ist. Zuletzt führen wir noch eine Schätzung durch, die berücksichtigt, dass ein Teil der Schüler, die von der Änderung der Pflichtschulregelungen betroffen waren, auch von einem Kurzschuljahr betroffen waren. Eine Rekodierung unserer endogenen Variablen, um die Kurzschuljahre zu berücksichtigen, führt ebenso zu keiner Veränderung unserer Ergebnisse.

Ein Zusammenhang zwischen Bildung und Gesundheit durch bessere Gesundheitsinputs im Sinne von weniger Rauchen wird durch unsere Daten nicht gestützt. Übergewicht und Fettleibigkeit können nicht nur als Indikatoren künftiger Gesundheitsprobleme sondern auch als Indikatoren des vergangenen gesundheitsbezogenen Verhaltens interpretiert werden. Daher stützen unsere Ergebnisse bezüglich des Körpergewichts indirekt das Argument der Gesundheitsinputs, wobei dies mehr für die Männer als für die Frauen gilt. Änderungen in der Beschäftigung von manuellen zu nicht-manuellen Tätigkeiten könnten eine Erklärung darstellen. In diesem Zusammenhang haben wir Einkommen und Löhne nicht untersucht. Aus der existierenden Literatur - unter Nutzung einer identischen Identifikationsstrategie - ergibt sich, dass die Veränderungen in der Anzahl der Pflichtschuljahre keinen kausalen Effekt auf die Löhne hatten (Pischke and von Wachter (2008)), so dass ein Zusammenhang zwischen Bildung und Gesundheit aufgrund eines höheren Einkommens (und daher günstigerer Lebensbedingungen) unwahrscheinlich erscheint. Schließlich soll noch darauf hingewiesen werden, dass unsere Parameter den Effekt der Bildung nur für diejenigen identifizieren, deren Verhalten durch die spezifischen Änderungen in den Pflichtschulregelungen beeinflusst wurde. Interventionen zu anderen Zeitpunkten innerhalb des Lebenszyklus oder zielgerichtetere Interventionen könnten stärkere und systematischere Effekte auf die Gesundheit und gesundheitsbezogenes Verhalten entfalten.

Kapitel 2: Das zweite Kapitel untersucht die Effekte der mütterlichen Gesundheit auf die Gesundheit sowie das gesundheitsbezogene Verhalten der Kinder. Eine Quantifizierung solcher intergenerationaler Zusammenhänge ist nicht nur für optimale Bildungsinvestitionen relevant sondern steht auch mit der sozialen Mobilität in Verbindung. Wir leisten einen Beitrag zur Literatur, indem wir einen IV Ansatz anwenden, der bei Nutzung von Haushaltsdatensätzen üblicher Größe funktioniert, indem wir eine Reihe von Gesundheitsmaßen sowohl für Neugeborene als auch für Jugendliche betrachten, und indem wir mögliche Kanäle der geschätzten Effekte untersuchen. Wir instrumentieren die mütterliche Bildung mit der Anzahl ihrer Geschwister, während wir für Charakteristika ihrer Eltern, d.h. der Großeltern des Kindes, kontrollieren. Wir greifen dafür auf einen reichhaltigen Haushaltsdatensatz, das deutsche sozio-ökonomische Panel, zurück, das detaillierte Informationen über drei Generationen enthält. Wir argumentieren, dass - gegeben die Charakteristika der Großeltern - die Anzahl der Geschwister der Mutter eine Variation in der mütterlichen Bildung identifiziert, die im Hinblick auf Gesundheit und gesundheitsbezogenes Verhalten ihrer Kinder exogen ist.

Für die Neugeborenen finden wir einen signifikanten negativen Zusammenhang zwischen der mütterlichen Bildung und der Wahrscheinlichkeit, dass es zu einer Frühgeburt kommt. Obgleich sich die Größe des geschätzten Effekts erhöht, wenn wir ein IV-Probit Modell schätzen, wird der Effekt insignifikant. Es könnte sein, dass unsere Stichprobe der Neugeborenen nicht groß genug ist, um einen existierenden Kausaleffekt nachzuweisen. Wir finden keine Evidenz für einen Effekt der mütterlichen Bildung auf die Wahrscheinlichkeit eines niedrigen Geburtsgewichts. Für Jugendliche weisen die IV Schätzungen auf starke und signifikante Effekte auf das gesundheitsbezogene Verhalten der Töchter hin. Ein zusätzliches Bildungsjahr reduziert - gemäß unserer Schätzungen - die Wahrscheinlichkeit, dass die Tochter im Alter von 18/19 Jahren raucht, um 7,4 Prozentpunkte und erhöht die Wahrscheinlichkeit, dass die Tochter mindestens einmal in der Woche Sport treibt, um 7,5 Prozentpunkte. Wir finden jedoch keine signifikanten Effekte der mütterlichen Bildung auf das gesundheitsbezogene Verhalten der Söhne. Für beide Geschlechter finden wir weder Effekte auf die physische Gesundheit noch auf die Wahrscheinlichkeit eines Übergewichts. Im Hinblick auf die Relevanz unseres Instruments, weisen alle F-Statistiken einen Wert auf, der größer als die kritische Marke von 10 ist.

Wir zeigen die Robustheit unserer IV Schätzungen, indem wir Charakteristika der Großeltern schrittweise in die Spezifikationen aufnehmen. Unsere Ergebnisse bleiben zudem unverändert, wenn wir für zusätzliche Variablen (wie die Gesundheit der Mutter, ihr gesundheitsbezogenes Verhalten, und ihre Fertilität) kontrollieren, die möglicherweise Folgen der mütterlichen Bildung sind.

Die Ergebnisse sind robust, wenn nur Mütter mit 1-3 Kindern betrachtet werden, sowie im Hinblick auf Annahmen der funktionalen Form und der Verteilung der Störterme. Bei einer Untersuchung möglicher Kanäle der geschätzten Effekte weisen unsere Ergebnisse weder darauf hin, dass gesundheitsbezogenes Verhalten der Mutter, noch dass selektive Partnerwahl oder das Haushaltseinkommen die Effekte auf das gesundheitsbezogene Verhalten der Töchter erklärt. Wenn wir jedoch den Bildungspfad der Tochter als eine zusätzliche Kontrollvariable im Rahmen unseres IV Ansatzes aufnehmen, verschwindet der Effekt auf das Raucherverhalten der Töchter vollständig. Die mütterliche Bildung scheint also das gesundheitsbezogene Verhalten der Töchter über den Bildungspfad zu beeinflussen.

Kapitel 3: Das dritte Kapitel enthält Evidenz für den Vorteil Bayesianischer Schätzverfahren anstelle der klassischen Maximum-Likelihood-Methode für die Schätzung dynamischer diskreter Entscheidungsmodelle. Diese Modelle erfordern üblicherweise eine allgemeine Spezifikation der unbeobachteten Präferenzheterogenität und es müssen daher oft relativ komplexe Algorithmen für die Schätzung angewendet werden. Wir schätzen ein dynamisches diskretes Entscheidungsmodell für das Arbeitsangebot von Frauen mit drei unterschiedlichen Zuständen und unterschiedlichen Spezifikationen für die unbeobachtete Heterogenität. Unsere Analyse basiert auf Längsschnittsdaten des deutschen sozio-ökonomischen Panels.

Der empirische Vergleich führt zu zwei wichtigen Schlussfolgerungen, die für die Anwendung sehr relevant sind, wenn dynamische diskrete Entscheidungsmodelle mit unbeobachteter Heterogenität geschätzt werden. Erstens, bei Annahme einer multivariaten Normalverteilung für die unbeobachtete Heterogenität führen beide Ansätze, der MCMC Schätzer und die MSL Schätzung, zu beinahe identischen Ergebnissen. Dies zeigt, dass bei einer endlichen Stichprobe mit einer Größe, wie sie für übliche Haushaltspaneldatensätze typisch ist, unsere Ergebnisse mit den asymptotischen Aussagen des Bernstein-von Mises Theorems in Einklang stehen. Die Ergebnisse der Bayesianischen Schätzung können also auch klassisch interpretiert werden. Das zweite Ergebnis zeigt den Vorteil des Bayesianischen Schätzverfahrens. Wir zeigen, dass unter Verteilungsannahmen, die mit der Theorie konsistent sind, z.B. log-normal verteilte Konsumpräferenzen, die Bayesianische Methode gut funktioniert und glaubwürdige Ergebnisse liefert, während der MSL Schätzer nicht konvergiert. Dies weist darauf hin, dass Bayesianische Schätzverfahren ein nützliches Werkzeug für die Schätzung dynamischer diskreter Entscheidungsmodelle sein können.

Kapitel 4: Das vierte Kapitel schlägt ein Lebenszyklusmodell vor, um die gesundheitsbezogenen Risiken des Konsums und der Altersarmut zu untersuchen. Zudem zeigt die Studie die hohe Leistungsfähigkeit einer Erweiterung des EM Algorithmus, indem ein komplexes dynamisches Modell der Gesundheitsrisiken, Arbeitsmarktpartizipation, Frühverrentungsentscheidung und Vermögensakkumulation geschätzt wird. Die Erweiterung des EM Algorithmus wird verwendet, um gute Startwerte für eine sich anschließende FIML Schätzung zu gewinnen. Ich stütze mich dabei auf den methodischen Rahmen eines DPDC Modells. Meine Analyse betrachtet männliche Singles und erlaubt es folglich Anpassungensreaktionen des Partners im Hinblick auf Arbeitsangebot und Renteneintrittsverhalten zu vernachlässigen. Die analysierten Mechanismen sind auch für Paare und weibliche Singles relevant. Die gesundheitsbezogenen Risiken sind vermutlich kleiner für Paarhaushalte, weil Anpassungsreaktionen des Partners die ökonomischen Konsequenzen einen Gesundheitsschocks abschwächen können. Es folgt daraus, dass gesundheitsbezogene Risiken männlicher Singles eine Obergrenze für die entsprechenden Risiken von Paarhaushalten anzeigen.

Ich simuliere Lebenszyklen von Individuen nach Anfangsausstattungen im Alter von 40 Jahren. Das Modell wird genutzt, um Szenarien zu simulieren, in denen Gesundheitsschocks zu unterschiedlichen Zeiten innerhalb des Lebenszyklus auftreten oder nicht auftreten. Ein Vergleich der Konsumpfade und Nettogegenwartswerte des erwarteten künftigen Konsums zwischen den Szenarien wirft Licht auf gesundheitsbezogene Konsum- und Armutsrisiken, die durch das deutsche Sozialversicherungssystem nicht versichert sind. Die Simulationen zeigen, dass erwartete gesundheitsbezogene Verluste hinsichtlich des künftigen Konsums substantiell von den Anfangsausstattungen abhängen und sich in der Bandbreite zwischen 3 % und 7 % bewegen. Die erwarteten Verluste sind größer für Individuen, die kein Nettovermögen im Alter von 40 Jahren aufweisen oder die ein niedrigeres Bildungsniveau haben. Zudem sind die erwarteten Verluste kleiner in Ostdeutschland als in Westdeutschland. Die Realisierung eines Gesundheitsrisikos führt zu hohen Verlusten im erwarteten künftigen Konsum, wenn der Gesundheitsschock zu einem frühen Zeitpunkt im Lebenszyklus auftritt. Die Ergebnisse weisen darauf hin, dass das deutsche Sozialversicherungssystem gesundheitsbezogene Konsumrisiken nicht hinreichend versichert. Selbstverständlich können Individuen zusätzliche private Berufsunfähigkeitsversicherungen erwerben. Diese Simulation zeigen lediglich die Grenzen des öffentlichen Sozialversicherungssystems auf.

Ich nutze die Simulationen des Lebenszyklus auch, um das gesundheitsbedingte Risiko der Altersarmut zu untersuchen. Die Größe des Risikos sowie gesundheitsbedingte Änderungen desselben hängen wesentlich von den Anfangsausstattungen eines Individuums ab. Das Risiko ist größer in Ostdeutschland als in Westdeutschland sowie für Individuen, denen es an Nettovermögen fehlt oder die Zeitperioden in Arbeitslosigkeit verbracht haben. Die gesundheitsbedingten Änderungen des Risikos sind substantiell für die betrachteten Anfangsausstattungen. Es gibt daher eine Problemstellung hinsichtlich gesundheitsbedingter Altersarmut, die durch das deutsche Sozialversicherungssystem nicht versichert ist. Aus diesem Grund untersuche ich ein Reformszenario für das deutsche Rentenversicherungssystem, das darauf abzielt, Individuen vor dem Risiko der Altersarmut zu schützen, indem ein Mindestrentenniveau auf der Höhe der Armutsgrenze eingeführt wird. Da eine solche Reform mit erhöhtem Missbrauch der Erwerbsminderungsrente und einem Rückgang des durchschnittlichen Renteneintrittsalters einhergehen könnte, untersuche ich, ob eine Bedürftigkeitsprüfung (d.h. Vermögensprüfung) den möglichen Missbrauchsanstieg reduzieren kann. Die Simulationen weisen nur auf einen geringen Rückgang des erwarteten Renteneintrittsalters hin (zwischen 0 und 0,4 Jahren in Abhängigkeit der Anfangsausstattungen), wenn die Gewährung der Mindestrente von einer Bedürftigkeitsprüfung abhängt. Ohne eine solche Prüfung ist der simulierte Rückgang des erwarteten Renteneintrittsalter substantiell (zwischen 0 und 3,7 Jahren). Im Großen und Ganzen lässt sich feststellen, dass eine Mindestrente mit Bedürftigkeitsprüfung einen gangbaren Weg darzustellen scheint, um Individuen vor dem Risiko der Altersarmut zu schützen.

Erklärung gem. §9(4) der Promotionsordnung der Freien Universität Berlin

Hiermit erkläre ich, dass ich meine Dissertation selbstständig verfasst habe.

Daniel Kemptner