Essays in Applied Econometrics and Health Economics

Inaugural-Dissertation
zur Erlangung des Grades
Doctor oeconomiae publicae (Dr. oec. publ.)
an der Ludwig-Maximilians-Universität München
vorgelegt von

Raphael Guber
2017

Referent:	Prof. Dr. Joachim K. Winter
Korreferent:	Prof. Axel Börsch-Supan, Ph.D.
Promotionsabschlussberatung:	31. Januar 2018

Tag der mündlichen Prüfung: 19.01.2018

Namen der Berichterstatter: Joachim Winter, Axel Börsch-Supan, Amelie Wuppermann

Acknowledgements

Out of the many people that accompanied me on this journey, I would first and foremost like to thank my supervisor and mentor Joachim Winter for his continuous support, effort, advice and honesty at any point of my dissertation. I am deeply grateful to my second advisor Axel Börsch-Supan for encouraging discussions and for providing me an ideal research environment at the Munich Center for the Economics of Aging, without which this dissertation would not have been possible. I thank Amelie Wuppermann for serving as a third reviewer on my dissertation committee.

Writing this dissertation was an intellectually stimulating and challenging experience, which I enjoyed very much. This is to no minor part due to my co-authors Helmut Farbmacher, Tabea Bucher-Koenen, Johan Vikstroem and Martin Kocher, whom I would like to thank for their critical remarks and advice. I would like to especially thank Helmut Farbmacher, Tabea Bucher-Koenen and Martin Spindler at MEA for their encouragement and support in the past years. I greatly enjoyed my research stay at the Harvard Center for Population and Development Studies and would like to thank Lisa Berkman for inviting me.

Probably one of the greatest experiences this dissertation offered was being part of the "Evidence-Based Economics" graduate program at the University of Munich. Being part of a cohort of highly motivated peers and, now, friends made this journey a enjoyable and memorable one. Here, I'd like to thank Joachim Winter and Florian Englmaier for their endless efforts in bringing this graduate program to a success for all its members. I am grateful to the Max-Planck Society and the Elitenetwork of Bavaria for providing funding at various occasions during my studies.

Finally, I thank Michaela for her loving support throughout the years.

Contents

List of Figures

List of Tables

Preface

Economics strives to gain insight in the determinants of human decision making and its consequences for both the individual and the aggregate economy. From this knowledge, the goal is to derive informed policy recommendations that improve the welfare of the society at large.

In light of the challenges posed by a rapidly changing nature of the labor market, the need for life-long learning, and an aging society (Börsch-Supan, 2003), a crucial relationship that requires a deeper understanding are the reciprocal effects of health and labor markets. Health is an established driver of individual's labor productivity and well-being (Smith, 1999). Both health and productivity are major determinants of economic growth and development (Bloom et al., 2004; Hanushek and Woessmann, 2008). Furthermore, technological progress and demographic change have led growth in public spending on health care to consistently outpace growth in GDP in rich countries (WHO, 2015). Preserving the sustainability of publicly funded health and other social insurances is a major challenge for many governments in the presence and future.

One important step in the process of generating knowledge for policy makers is to provide data-based evidence on the causal impact of reforms, interventions and individual choices. Isolating causes from correlates is no easy task. It requires the precise manipulation of one factor of interest while holding all other factors constant. This is rather easily possible in theoretical models, but in the real world, the factor of interest is connected to a multitude of other variables that move together simultaneously, thereby blurring the causal effect. Significant progress has been made on the methodological possibilities for the identification of causal effects in observational data in recent years, which have considerably increased the credibility of empirical findings (Angrist and Pischke, 2010).

These methods build on the exploitation of natural experiments. In contrast to e.g. laboratory experiments, which provide the researcher with a controlled environment, in natural experiments an event outside of the researchers control generates quasi-experimental variation in treatments (Rosenzweig and Wolpin, 2000). This approach has now become a standard tool in applied economic research using observational data (Angrist and Pischke, 2008). Indeed, some research questions can only be answered by natural experiments, because randomized controlled trials would be too expensive, unethical or politically impossible, such as randomly sending some individuals to war to study the effect of veteran status on labor income (Angrist, 1990).

The need to inform policy debates on the relationship of health and labor markets and the development of econometric methods to effectively do so is the starting point of my thesis. Its aim is to provide credible, data-based insights. More specifically, it contributes to the knowledge on how early childhood interventions influence adult labor productivity, how fertility decisions affect female labor market attachment and, in interaction with the latter, mortality in old age.

This thesis consists of four self-contained chapters. A common theme across all chapters is the application and advancement of methods that estimate causal effects. In the first chapter, I apply a difference-in-difference strategy to estimate the causal effect of forced right-hand writing of left-handers on labor market outcomes. The second chapter proposes a method to eliminate sample selection bias in the twin birth instrument, which is often used to instrument fertility decisions in various settings. The third chapter relies on reduced form estimation to study the long-run consequences of fertility and its interaction with labor market participation on mothers' mortality in old age. While the previous chapters are based on natural experiments, the fourth chapter employs a laboratory experiment to investigate whether insuring against losses makes individuals overconfident in their own ability relative to others. The four chapters are now motivated and described in more detail.

In the first chapter, I investigate whether early childhood interventions can compensate or engrave innate disadvantages. This essay is motivated by work of James Heckman and his co-authors who in a series of papers have shown that early childhood interventions, such as educational support and parental counseling, can have large long-run benefits for socio-economically disadvantaged children (Heckman, 2000; Cunha and Heckman, 2010). I study whether the forced right-hand writing ("switching") of children who were born as left-handers has similar consequences for labor market outcomes later in life. Left-handers are a particular interesting population in this context, as previous research has shown that, on average, left-handed children perform significantly worse in standardized math and reading tests, obtain fewer years of schooling, and are more likely to suffer from learning disabilities and behavioral problems (Goodman, 2014). Left-handed adults earn lower wages because they select into occupations that require lower levels of cognitive skills.

Using data from the German Socio-Economic Panel (GSOEP), I find that switched left-handers perform equally well on the labor market and have similar levels of education than right-handers, the control group, while non-switched left-handers experience the previously documented deficits associated with left-handedness. I show that non-cognitive skills, personality traits and locus of control, also differ between these three groups, but do not explain the gap observed among labor market outcomes. Schooling stands out as the single most important mediating variable between switching and wages. These results are robust to various model specifications as well as sample and treatment status definitions.

Nevertheless, unobserved variables such as parental endowment and the child's motivation might be the true cause behind switching and the outcome variables. To address potential selection bias, I employ a difference-in-differences approach, where I exploit the variation in switching across cohorts and handedness. In effect, cohort trends of the outcome variables of right-handers, who were never switched, are used as a counterfactual for left-handers, thus allowing for a time-constant difference between left- and right handers. I argue that this strategy will deliver downward biased estimates on the causal effect of switching, but observe that OLS and IV estimates differ only little from each other.

I conclude that, even though potentially harmful for the child in the short run, switching has had a positive effect on adult outcomes, compensating for innate deficits. One potential explanation is that these children receive additional attention and care by their teachers and parents, resulting in higher human capital accumulation. A successful switching of the writing hand may also induce a feeling of success and motivate children to improve their skills further in the future. This is in line with Heckman's conclusion that motivation, more than cognitive skills, is a crucial determinant of childrens' future success.

The second chapter is joint work with Helmut Farbmacher and Johan Vikstroem. There exists a huge interest in the causal effect of fertility on mothers' economic outcomes, in particular labor market participation (Gronau, 1973). However, a simple comparison between mothers with different numbers of children does not deliver a causal effect, as mothers differ also in other, potentially unobservable dimensions, e.g. a preference for having a career, which could drive the observed relation between fertility and employment. Therefore, most papers use instrumental variable (IV) techniques. One commonly employed instrument are twin births (Angrist and Evans, 1998; Mogstad and Wiswall, 2016; Lundborg et al., 2017). The birth of twins is a natural experiment, which randomly increases some mothers' number of children, independent of their unobserved preferences. However, it has been questioned if having twins really is a random event. In particular, dizygotic (fraternal) twinning depends on, for example, maternal age, height, weight, race, and the use of fertility treatments, such as in-vitro fertilization (IVF). On the other hand, monozygotic (identical) twin births are considered a random event, since they are the result of the random and spontaneous division of a single fertilized egg.

In this essay, we propose a new instrument based on monozygotic twin births which corrects for the non-randomness of dizygotic twin births. Our key assumption is that monozygotic twinning is exogenous, but since zygosity is rarely known our approach does not rely on observing zygosity directly. We show that it is possible to use the observed opposite-sex dizygotic twin mothers to correct the same-sex twin instrument by the remaining selection bias induced from the same-sex dizygotic twins. This is possible because we know that all monozygotic twins are of the same-sex and that dizygotic twin births with same-sex twins are equally likely as dizygotic twins with opposite-sex. Our new approach can easily be implemented using standard regression techniques.

The new instrument is applied to US Census data and Swedish administrative data. In line with our expectations, we find that the new instrument delivers larger estimates on the negative effect of children on maternal labor market outcomes than the previous twin instruments. This is because mother who get dizygotic twins are a positively selected group, a fact that we also demonstrate using our administrative data.

Our newly developed instrument is a strong improvement over existing instruments for fertility decisions. As fertility treatments, in particular in-vitro fertilization, become more common among younger cohorts, the original twin instrument, which assumes randomness of twinning, becomes less credible over time. Restricting data sets to older cohorts of mothers were this issue is less relevant is an unsatisfying alternative. We thus believe that our new instrument is highly relevant and even necessary for future research that attempts to study the causal effect of fertility decisions on various outcomes.

The third chapter is joint work with Helmut Farbmacher, Tabea Bucher-Koenen and Johan Vikstroem. As female labor market participation rates increase, the potential double burden posed by raising children and having a career and its effect on maternal health in old age becomes of significant research interest (Sabbath et al., 2015). Actively raising children and pursuing a career are two conflicting, because time consuming, activities. One often made argument in the public debate on how to improve the reconciliation of family and working life is that mothers (and fathers) need to be shielded from stress. However, there exists actually very little evidence on the existence and long-run effects of this double burden on maternal health (Cáceres-Delpiano and Simonsen, 2012; Kruk and Reinhold, 2014). This essay aims at filling this gap in the literature, focusing on maternal mortality in old age.

The analysis is based on linked administrative birth and death registries from Sweden which enable us to reliably link children to their parents. The sample for our analysis includes more than 400,000 mothers that were 55-65 years old, alive and resident in Sweden in 1990. We can follow these mothers over time for twenty years until 2010. Since we cannot directly measure life-time stress, we analyze mortality due to two specific groups of medical diagnoses that have been related to stress during life in the literature: Cardiovascular diseases, specifically heart attacks and strokes, and smoking-related diseases, specifically lung cancer and chronic obstructive pulmonary disease (COPD).

In order to study the combined effect of fertility and labor force activity on maternal health, an ideal set-up would provide exogenous variation in both labor market participation and having and raising children. We use twins at first birth as an unplanned shock to fertility. While previous studies used the birth of twins as an instrumental variable (IV) for fertility, we study the reduced form effects of twinning and interpret them as being caused by a random event. The reason for this approach is that, in the context of health, the birth of twins might violate the exclusion restriction. The issue of non-random twinning discussed in the second chapter of my thesis is much less of a problem here, as the mothers in our sample had their first birth between 1940 and 1970, well before the availability of in-vitro fertilization.

In order to find variation in labor force attachment we stratify the sample along two variables which are strongly related to labor market activity, educational attainment and pension income at age 70. One obvious worry when splitting the sample in this way is selective sorting. Twins at first birth could directly affect the level of education, pension income or survival and retirement until age 70. We show that these concerns do not realize in our data.

The approach in this paper is a significant improvement on past research that studies the interacting effects of fertility and working life and maternal mortality. Existing research investigated differences in mortality rates across groups of mothers characterized by stylized work-family profiles, but was not able to control for selection into these profiles (Sabbath et al., 2015).

We find that mothers' probability to die over a period of 20 years is strongly increased when having twins at first birth. Moreover, the effects are largest among highly educated mothers and those with above-median pension income. These results are is in line with our hypothesized double burden effect.

Our findings have important policy implications. Excess deaths due to the described double burden have to be considered in the cost-benefit analysis of future family-friendly policy measures. Furthermore, our findings with respect to mortality from stress-related diseases hint at increased costs for the health care system over the adult life course. Policies that aim at alleviating stress from raising children and pursuing a career could help in avoiding such long-run costs. Our findings are particularly important as among younger generations, an increasingly larger group of women stays attached to the labor force until old age (Goldin and Mitchell, 2017).

The fourth chapter is joint work with Joachim Winter and Martin Kocher. We start from the observation that overconfidence, as other behavioral biases, has now been established as an important dimension and driver of individual's economic behavior (Thaler, 2000). Behavioral biases have a significant impact on contract design in many settings. For example, overconfidence has been found to predict excess market entry of entrepreneurs, risky investment decisions of CEOs and speculative trading (Camerer and Lovallo, 1999; Scheinkman and Xiong, 2003; Malmendier and Tate, 2005). In the context of insurance, it has been shown that if the share of overconfident individuals in the population is large enough, compulsory insurance is not pareto optimal anymore (Sandroni and Squintani, 2007).

However, this literature takes over- or underconfidence in individuals implicitly as pre-determined or fixed. This is not unique to confidence however, as other behavioral aspects, such as loss aversion or present bias, are equally assumed to be stable within individuals in a certain decision environment. This paper provides evidence for self-confidence to be malleable in a setting that has relevant implications. We show in a laboratory experiment that confidence in one's own performance depends on whether people acquire insurance or not.

More specifically, we develop an experimental design that allows us to cleanly disentangle effects from the incentives provided by the insurance contract from effects coming from selection into the contract. In our setup, an insurance partially covers potential losses from bad performance in a real-effort task. Before solving the task, individuals are given the choice to buy an insurance contract. Conditional on this choice, actual insurance status is randomized. The first part allows us to measure pure selection effects, while the second part identifies pure incentive effects.

Our results are consistent with insurance increasing individual's confidence as compared to a control group. At the same time, we find no evidence for more confident individuals choosing more or less insurance in the first place.

These findings have important implications for the design and research on insurance contracts. Instead of focusing only on issues resulting from self-selection based on over- or under-confidence, researchers should take incentive effects into account and design contracts to counter the increase in overconfidence, which is known to correlate with risky behavior.

Chapter 1

Making it right? Left-handedness, Social Norms and Human Capital

Abstract: *Can early childhood interventions compensate for innate deficits? In this paper, I study the forced right-hand writing of left-handed children ("switching"). While previous literature has found that, due to innate cognitive deficits, left-handers obtain less human capital and lower wages than right-handers, I find that switched left-handers perform equally well or even better in the labor market than right-handers. Only non-switched left-handers exhibit the deficits of left-handers found in earlier studies. To address potential selection bias, I employ a difference-in-difference approach, where I exploit the rapid decline of switching across cohorts. Cohort trends of the outcome variables of right-handers, who were never switched, are used as a counterfactual for left-handers. Using rich data from the German Socio-Economic Panel (SOEP), I show that the observed differences in outcomes occur due to differential human capital accumulation, rather than cognitive or non-cognitive skills. My findings are consistent with switching compensating for the innate deficits of left-handers.*

1.1 Introduction

Experimental evidence demonstrates that early childhood interventions at school entry age are followed by huge benefits later in life (Cunha et al., 2006; Heckman et al., 2013). Since then, economists and policy makers have become increasingly interested in ways to apply such interventions to the general population (Cunha and Heckman, 2010). However, there exist few studies that exploit naturally occurring interventions and that are able to look at long-term effects, see Currie and Almond (2011) for an overview.

In this paper, I study forced right-hand writing of left-handers, called switching from now on, as a case where parents invest into their children at an early age. I analyze the long-run consequences on labor market outcomes in adulthood and investigate a set of potential channels, ranging from human capital accumulation to cognitive and non-cognitive skills. Forced right-hand writing is motivated by stigma against left-handedness which varies by cohort.

To address potential selection bias, I employ a difference-in-differences approach, exploiting the variation in switching across cohorts and handedness. In effect, cohort trends of the outcome variables of right-handers, who were never switched, are used as a counterfactual for left-handers, thus allowing for a time-constant difference between left- and right handers. I argue that this strategy will deliver biased downwards estimates on the causal effect of switching, but observe that OLS and IV point estimates differ only little from each other.

Left-handers are a particularly interesting population for an early childhood intervention, as they also significantly differ from right-handers with respect to cognitive and non-cognitive skills caused by different brain structures.[1] Recent literature finds that, on average, left-handers experience deficits in skills and human capital accumulation when compared to right-handers. Using five comprehensive data-sets from the US and the UK, Goodman (2014) shows that left-handed chil-

[1] The importance of these traits for long-term economic performance is studied by Borghans et al. (2008).

dren perform significantly worse in standardized math and reading tests, obtain fewer years of schooling, and are more likely to suffer from learning disabilities and behavioral problems. Left-handed adults earn lower wages because they select into occupations that require lower levels of cognitive skills. Johnston et al. (2009, 2013) find that left-handed children in the Longitudinal Study of Australian Children (LSAC) and in the National Longitudinal Survey of Youth (NLSY) perform worse on cognitive development test scores than right-handed children. On the other hand, Denny and O'Sullivan (2007) find a wage premium for left-handedness among males and a wage penalty for women in the National Child Development Study (NCDS), see also Ruebeck et al. (2007). The child samples in the above studies are drawn from countries and cohorts in which switching and stigma against left-handedness is rather rare. Thus, they demonstrate a natural difference between left- and right-handers at a young age.

Whether switching increases or compensates for such innate deficits is highly informative for other early childhood interventions which also target vulnerable populations. As far as the knowledge of the author extends, no study concerning a non-institutional intervention, with the potential to have negative effects on those treated, has so far appeared in the literature.

Surprisingly, I find that switched left-handers perform equally well or even better than right-handers in terms of labor market outcomes and human capital accumulation, while non-switched left-handers exhibit the previously documented deficits of lefties. Cognitive skills, which are measured at adulthood, differ little, while non-cognitive skills are significantly different between left-(switched and non-switched) and right-handers. However, these differences explain only a small part of the observed gaps in labor market performance. The most important channel is human capital. Taking into account human capital accumulation, switched left-handers show about the same wage deficits as non-switched left-handers. These findings are consistent with switching as a compensatory investment for the innate deficits of left-handers.

My data come from the German Socio-Economic Panel (SOEP), a large and representative panel survey of the German population that provides a unique opportunity to observe this intervention. The data set covers a wide range of cohorts, individuals born between 1920 and 1997, with considerable variation in the prevalence of switching. Starting among cohorts born in 1950, switching rates decline monotonically from about 90% to 60% by 1960 and to nearly zero in 1990. Across all cohorts, 57% of left-handers are switched. In contrast, the share of naturally born left-handers remains fairly constant, at 9%, in particular from the 1940 cohort on.

This paper proceeds as follows. In the next section, I briefly review some of the literature on left-handedness and switching. Section 1.3 introduces the data and gives a descriptive analysis of left-handedness and switching across cohorts. Section 1.4 sets up a simple parental investment model that accommodates social norms to predict which parents are likely to switch their child. These predictions are then tested empirically. In Section 1.5 I present the differences between right-handers and switched and non-switched left-handers in labor market performance and discuss potential channels. In Section 1.6 I outline a strategy to identify the causal effect of switching. Finally, I conclude in Section 1.7.

1.2 Handedness and switching

Left-handers have faced discrimination in various areas of life (Harris, 1980, 1990). To a significant extent, prejudices about left-handers' inferiority have originated in religion, but not exclusively so. For example, in Christianity, the left hand was considered to be the devil's hand, and in Islam it is forbidden to use the left hand for eating and human interaction. Nonetheless, non-religious China has one of the lowest reported left-handedness rates worldwide, where right-hand writing is a social convention, rooted in the stroke order of Chinese characters (Kushner, 2013). Such attitudes may explain why switching the writing hand of left-handers is still common in developing countries in Asia, Africa, the Middle East, and South America (Medland et al., 2004; Zverev, 2006; Porac and Martin, 2007; Kushner, 2013). On more practical grounds, the world is primarily designed for right-handers, in particular, machinery, equipment and tools in everyday use. For these reasons, forcing a left-handed child to use the right hand for writing and other daily activities seems to be to the child's long-term benefit. Today, however, parents and teachers are advised not to interfere with a child's natural handedness, as it can lead to stuttering (Sattler, 1996).[2]

The share of left-handers in the population is estimated at 10%–15%, with variation in the country and cohort under study (Perelle and Ehrman, 2005).[3] A large literature on left-handedness, or laterality, exists in neuro-psychology, the neuro-sciences, and related fields (Coren, 2012). The origins of left-handedness are still unclear, however. Recent large-scale twin studies have shown that early theories based on a simple genetic model cannot be sustained (McManus et al., 2013). Satz (1972) proposes the idea of a pathological left-hander. According to this theory, even mild damage to the left brain hemisphere during the pre- or perinatal period can cause a shift of lateral dominance to the right hemisphere. Hence, lower cognitive skills, behavioral problems and left-handedness have the same cause. Goodman (2014) concludes that his findings are in line with the idea of a pathological left-hander.

In contrast to handedness, the consequences of switching are much less well researched, whether in psychology or any other field.[4] Depending on the country and cohort considered, the methods of switching range from friendly persuasion and positive incentives to threats, parental neglect, immobilization, beatings, and even breaking the left-arm (Perelle and Ehrman, 1994; Zverev, 2006). What consequences of switching can be expected? Sattler (1996) reports that in school,

[2]King George VI ("The King's Speech") is a well-known example from generations of left-handers affected in this way (Kushner, 2011, 2012).

[3]See McManus (2009) for an overview of the prevalence of left-handedness across time and geography.

[4]Previous work in psychology includes, e.g., Porac et al. (1986); Porac and Searleman (2002).

children forced to switch have to invest an over-proportional share of their energy and concentration in learning to write with the wrong hand. Hence they are quickly exhausted and are less able to follow the lessons.

Switching may also alter the brain's structure. Klöppel et al. (2010) find that the volume of gray matter in the putamen, a part of the forebrain that contains the executive and cognitive aspects of motor control, is reduced among switched left-handers, compared to non-switched right- and left-handers. A positive effect of switching might stem from increasing the brain's connectivity, such as the corpus callosum, which is known to be larger among left-handers (Witelson, 1985).

To summarize, some aspects of switching suggest negative consequences, either via physiological (overtaxing of the non-dominant brain hemisphere) or psychological (social exclusion, violence) channels, while others could have a positive effect, such as stimulating brain activity and additional attention from parents and teachers.

1.3 Data and descriptives

1.3.1 Left-handedness and switching

My sample is drawn from the German Socio-Economic Panel (SOEP), a large and representative panel survey of the German population that was started in 1984 (Wagner et al., 2007). As part of a grip strength measurement module conducted biannually between 2006 to 2014, respondents were asked "Are you a natural right- or left-hander?" and "With which hand do you actually write?".[5] I take a difference in the answers to these questions as an indication for switching of the writing hand. In this study, an individual is defined as being naturally left-handed if he reports

[5]In 2006 and 2014, individuals had "left- and right-hander" as a third answer option for both questions. I assign these 28 (13 for writing) individuals to the left-handed group (the writing hand is the left hand). Qualitatively, the findings do not change if I include them in the right-handed group. See Ambrasat and Schupp, 2011 (in German only) for further details on the grip strength measurement.

so at least once in any wave of the SOEP. The reasoning for this approach is that no true right-hander has an incentive to ever report being left-handed. Similarly, an individual is defined as switched if she reports a difference between her innate and writing hand during the same interview at least once. Reassuringly, only 32 out of a gross sample of 13,442 individuals report being innately right-handed, but write with their left hand today. I drop these observations.

The analysis sample is restricted to individuals born after 1920 and before 1997, in order to avoid small cell sizes. The total sample size is 12,757, of which 1,129 observations are left-handed and 633 switched. The resulting share of left-handers is 8.85%, of which 56.06% are switched. This share of left-handers is lower than the 10% to 15% reported in recent economic (Goodman, 2014; Johnston et al., 2013) or psychological (McManus, 2009) studies. One explanation for this is that these studies are able to create a more precise measure of handedness by combining statements on the preferred hand for writing, throwing, and eating, while my data set provides only one item on this trait. I now discuss the limitations of my measures.

One obvious concern is different reporting behavior by true handedness, true switching status and across cohorts. The way respondents interpret the term "natural right- or left-hander" may be directly influenced by these factors. Given the stigma of left-handedness, social desirability might lead to underreporting left-handedness or to missing values on the handedness questions. I use the left-to-right grip strength ratio to check whether true left-handers are less likely to answer the questions on innate handedness in the first place, but find no evidence for this.[6]

[6]Information on innate and writing hand is available for more than half of the individuals who refused the grip strength measurement. Similarly, for more than half of the individuals with missing information on handedness, information on grip strength measures are available. I compare the distributions of the left-to-right ratio of grip strength between those with and without missing information on handedness. The reasoning is that in my data, left-handers left-to-right grip strength ratio is on average more than one-third of a standard deviation higher than that of right-handers. A Wilcoxon–Mann–Whitney test does not reject the hypothesis that the two distributions (missing vs. non-missing information on innate handedness) of the grip strength ratio are equal (p-value 0.43), suggesting that left-handers are not over-represented among those with missing handedness information.

Furthermore, if the switching practice disappears across cohorts, we would expect that the share of reported left-handers increases, since switching that occurred too early in life to be remembered decreases. As I will show below, this is not the case.

Another concern may be that older individuals are less likely to write at all and hence report being right-handed. I offer two pieces of evidence against this hypothesis. First, Figure 1.1 shows the share of non-response to the original handedness and writing hand questions. I find no evidence that certain cohorts are more or less likely to answer the questions. Linear or quadratic cohort trends are non-significant for either question. Second, in unreported fixed effects logit regressions, I make use of the panel structure of the grip strength measurement module and find that age is not a significant predictor of the writing hand. Nevertheless, in robustness checks, I adopt different assignment rules for left-handedness and switching and find that my results still hold.

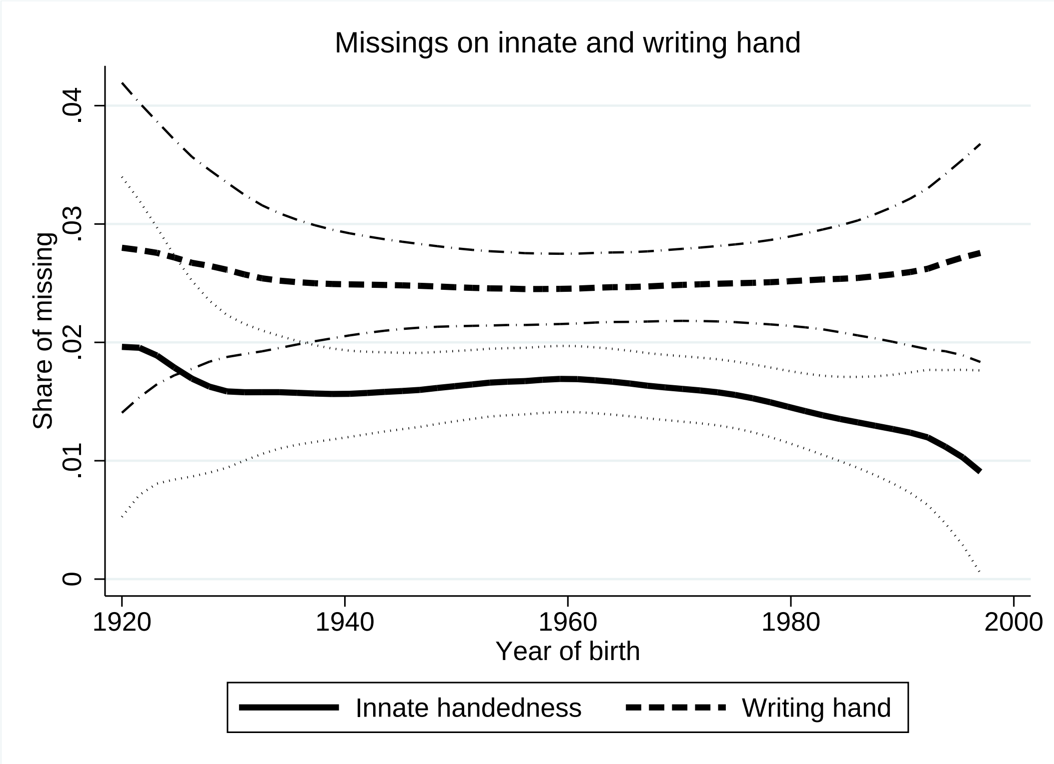
1.3.2 Outcome, channel and control variables

Employment status and log-hourly wages observed between 2004 to 2014 are the primary outcomes of interest in this study. Labor market outcomes can be seen as a summary of long-term consequences from early interventions (Heckman et al., 2013). I then investigate human capital, cognitive and non-cognitive skills as potential channels of switching as an intervention, from early childhood to labor market outcomes (Heckman et al., 2006).

Channels: Human capital Human capital accumulation is measured by completed years of education and retrospective grades in Math and German from the last school certificate. Years of education includes time spent in apprenticeships, training, or tertiary education. I use the highest observed value for years of education in the panel, but individuals had to be at least 25 years old at this point to be included in the sample.

The school is an integral part in the switching process, as it is the primary place

Figure 1.1: Share of missing observations



Note: Share of missing observations on handedness and writing hand by year of birth. Includes 95% confidence intervals.

to develop writing skills. Teachers in Germany have considerable discretionary power over tracking, which starts very early, after elementary school at age 10. Only the highest track (*Gymnasium*) leads to college education, see Krueger and Pischke (1995) for a more detailed description. The East German schooling system had no such early tracking, but access to the higher education granting track was strongly limited and required alignment with the state's ideology, see Baker et al. (2007).

Math grades serve as the earliest available proxy for cognitive skills, before college or occupational choices are made. In addition, writing and verbal skills, which might be influenced by switching, are much less important in Math than in German. Individuals are only included in the sample of grades if they were at least of age 20 at the time of the interview and thus likely to have completed schooling. Note, however, that grades are self-reported and therefore subject to recall bias, which may differ by handedness or writing hand. For example, left-handers might report worse grades because their memories of schooling is tainted by discrimination they experienced.

Channels: Cognitive skills For cognitive skills, I use the symbol-digit test (SDT) and the animal naming task (ANT).[7] Both were elicited in 2006 and 2012 from a random sample of SOEP participants and have been used in other studies (Dohmen et al., 2010; Heineck and Anger, 2010). During the SDT, individuals had to match as many numbers to symbols as possible within 90 seconds and enter their answers in the interviewer's computer. This test intends to measure an individual's fluid intelligence, which is the ability to process and make use of new information that is not already stored in the memory (Cattell, 1987). The ANT is a mixture of fluid (word fluency) and crystallized (vocabulary) intelligence. It requires respondents to name as many distinct animals as possible in 90 seconds. I use the number of uniquely named animals, excluding repetitions.

Channels: Non-cognitive skills Cunha et al. (2006) document that IQ gains in the Perry Preschool Program were short-lived and faded out within six years after the intervention. In contrast, non-cognitive skills such as self-motivation were responsible for the program's positive effect on later life outcomes. In this study, non-cognitive skills are represented by the Big Five personality traits (McCrae and Costa Jr, 1999) and the external locus of control (Rotter, 1966). Personality traits were elicited in 2005, 2009, and 2013, using three items for each trait.[8] The external locus of control was elicited in 2005 and 2010 using six items. I do not use all items in constructing these variables, due to their low reliability. See Appendix B for an overview on the construction of these variables. The external

[7]See Lang et al. (2007) for the validity and reliability of these tests in the SOEP.
[8]See Dehne and Schupp (2007) for the validity and reliability of these measures in the SOEP.

locus of control has been found to be an important predictor of wages (Groves, 2005; Heineck and Anger, 2010) and job search strategies (Caliendo et al., 2015). The role of personality traits for labor market returns in the SOEP is analyzed in Heineck and Anger (2010). For both, cognitive and non-cognitive channels, I use the earliest available observation per individual.

Control variables My control variables comprise gender, year of birth, being born in East Germany (the former German Democratic Republic)[9], migration background (none, 1st generation, 2nd generation), mother's and father's education (none/basic, middle, high, and missing) and urbanization at age 15 (large city, mid-sized city, small town, countryside, and missing). Individuals whose country of birth is not Germany are referred to as migrants.

Table 1.1 shows descriptive statistics by handedness and switching status. Simple comparisons of the means show that left-handers report worse grades in German and score lower on conscientiousness and extraversion than right-handers. Left-handers are less likely to be female, East German, or non-native. A more detailed analysis of reported left-handedness is deferred to Appendix 1.B. The fact that females are less likely to report left-handedness than males is well-known in the laterality literature (Harris, 1990) and I find no evidence that this difference changes across cohorts, suggesting a true biological cause. Differences by country of birth and between East and West Germany can be explained by past and prevailing anti-left attitudes, which are discussed in the Appendix. Apart from these basic characteristics, there are no significant differences in socio-economic background. Unconditional mean comparisons between switched and non-switched left-handers are not very informative here, as they are strongly confounded with cohort effects, which will be corrected for in regression analysis. Selection into switching will be discussed in Section 1.4.

[9] East German is defined by having lived in the GDR in 1989 or by being in sample C (D-Ost) in the SOEP.

Table 1.1: Descriptive statistics

	All		Left-handers			
	Right	Left	Non-switched	Switched	Total	N
Share in sample:	91.15%	8.85%	43.93%	56.07%		
Employed	0.821	0.812	0.775	0.839	0.820	53,213
	(0.383)	(0.391)	(0.417)	(0.367)	(0.384)	
Log-hourly wage	2.597	2.589	2.532	2.629	2.596	43,514
	(0.639)	(0.661)	(0.668)	(0.653)	(0.641)	
Years of Education	12.350	12.386	12.321	12.425	12.353	11,249
	(2.769)	(2.874)	(2.961)	(2.822)	(2.778)	
Math grade	0.005	-0.053	-0.245	0.078	0.000	7,541
	(1.002)	(0.981)	(1.067)	(0.896)	(1.000)	
German grade	0.006	-0.062	-0.158	0.004	0.000	7,265
	(1.001)	(0.990)	(1.035)	(0.955)	(1.000)	
Higher track	0.608	0.585	0.617	0.566	0.606	9,940
Symbol-Digit Test	-0.005	0.041	0.448	-0.214	0.000	5,033
	(0.993)	(1.061)	(1.021)	(1.006)	(1.000)	
Animal Naming Test	0.007	-0.071	0.027	-0.123	0.000	2,275
	(0.994)	(1.052)	(1.078)	(1.039)	(1.000)	
Openness	-0.003	0.031	0.076	0.001	0.000	11,037
	(1.002)	(0.974)	(0.966)	(0.979)	(1.000)	
Conscientiousness	0.006	-0.059	-0.131	-0.011	0.000	11,032
	(1.000)	(0.997)	(1.051)	(0.958)	(1.000)	
Extraversion	0.007	-0.070	-0.007	-0.112	0.000	11,038
	(0.999)	(1.009)	(0.995)	(1.017)	(1.000)	
Agreeableness	0.005	-0.049	-0.054	-0.045	0.000	11,041
	(1.000)	(0.997)	(1.011)	(0.988)	(1.000)	
Neuroticism	-0.004	0.036	-0.018	0.072	0.000	11,041
	(1.000)	(1.003)	(1.019)	(0.992)	(1.000)	
External locus of control	0.000	-0.000	0.026	-0.015	0.000	8,037
	(1.002)	(0.980)	(0.931)	(1.008)	(1.000)	
Year of birth	1962.2	1962.9	1975.2	1953.2	1962.2	12,757
	(18.949	(18.495)	(14.997)	(14.848)	(18.909)	
Female	0.533	0.477	0.488	0.469	0.528	12,757
East German	0.220	0.192	0.115	0.253	0.218	12,757
Migration background						
None/Native	0.791	0.827	0.788	0.858	0.794	12,757
1st Generation	0.090	0.093	0.123	0.070	0.090	12,757
2nd Generation	0.119	0.080	0.089	0.073	0.116	12,757
Father's education						
None/Basic	0.604	0.614	0.510	0.695	0.605	12,757
Middle	0.170	0.163	0.198	0.136	0.169	12,757
High	0.132	0.139	0.192	0.098	0.133	12,757
Missing	0.094	0.084	0.100	0.071	0.093	12,757
Mother's education						
None/Basic	0.632	0.637	0.524	0.725	0.633	12,757
Middle	0.211	0.205	0.258	0.163	0.210	12,757
High	0.087	0.096	0.147	0.055	0.088	12,757
Missing	0.070	0.063	0.071	0.057	0.069	12,757
Urbanization at age 15						
Large city	0.203	0.206	0.198	0.213	0.204	12,757
Mid-size city	0.168	0.201	0.226	0.182	0.171	12,757
Small town	0.230	0.224	0.258	0.197	0.230	12,757
Rural	0.380	0.352	0.292	0.398	0.377	12,757
Missing	0.019	0.017	0.026	0.009	0.019	12,757
Religious affiliation						
Catholic	0.313	0.321	0.314	0.326	0.313	10,189
Protestant	0.359	0.374	0.415	0.349	0.360	10,189
None	0.274	0.259	0.210	0.289	0.272	10,189
Other	0.055	0.046	0.061	0.037	0.054	10,189

Table displays means and standard deviations of non-binary variables in parenthesis below. Left-handed equals one if a respondent in the German SOEP reports at least once to be a natural left-hander during grip strength measurements performed in 2006, 2008, 2010, 2012 and 2014. Switched equals one if respondent reports at least once a difference between her natural and writing in the same year. All switched individuals are left-handers. Sample restricted to cohorts born between 1920 and 1997. SDT is the sum of correct entries in the symbol digit task within 90 seconds. ANT (90s) refers to the sum of uniquely named animals in the animal naming task within 90 seconds. SDT, ANT, grades, Big Five personality traits, and locus of control are standardized in full sample. Employment status and log(wage) applies for individuals observed at age 25 and 60 between years 2004 to 2013.

1.3.3 Left-handedness and switching across cohorts

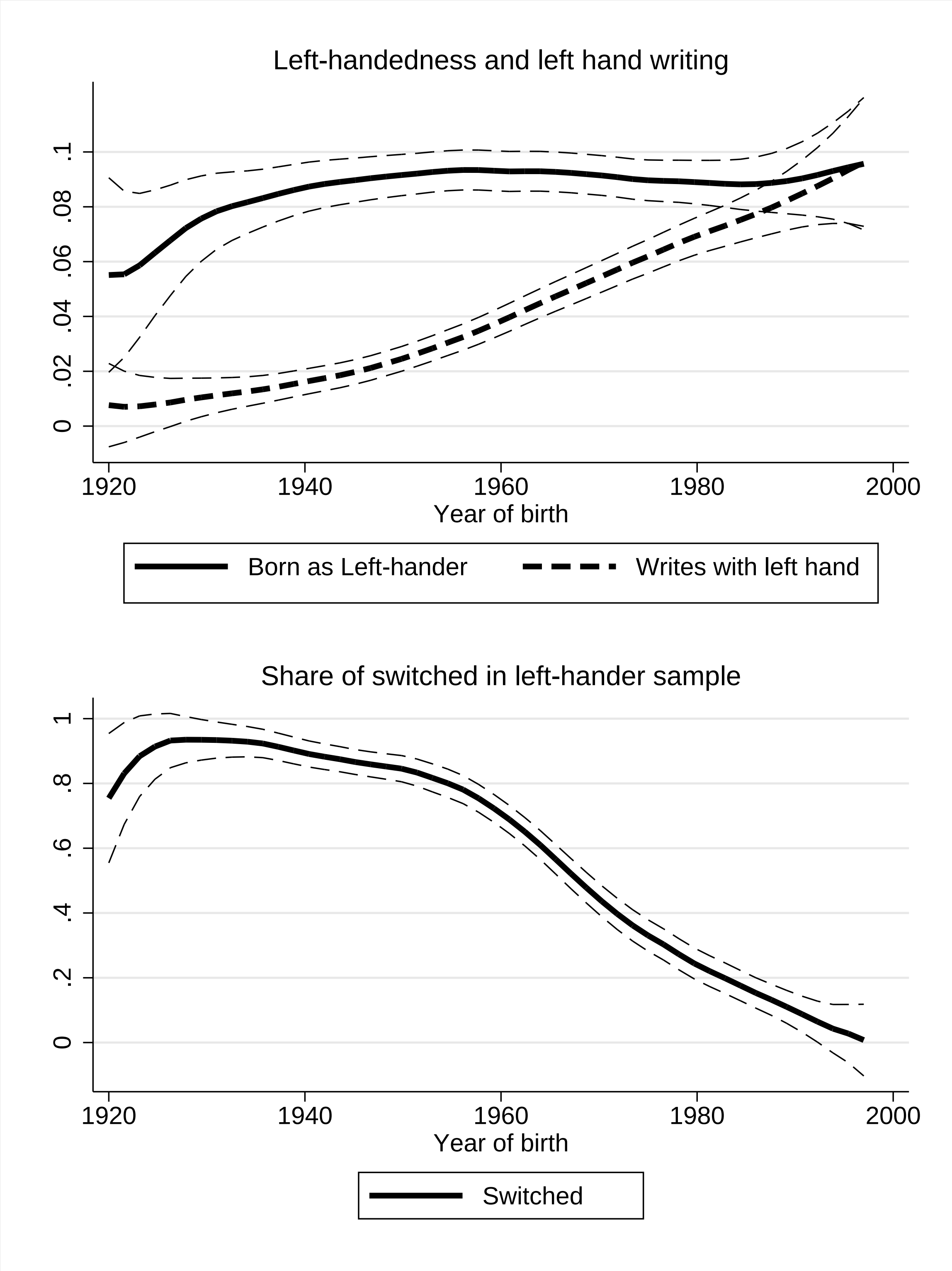
This section provides a description of switching and left-handedness in the data. The first panel in Figure 1.2 shows the share of innately left-handed individuals and left-hand writers across cohorts. The second panel only uses left-handers and shows the share of switched across cohorts. All curves are estimated by local linear regression with ROT bandwidth. Starting with the first graph, I find that the share of left-handers increases, though rather noisily, from 6% in the 1920 cohort to about 9% in 1940. It remains constant from then on. The initially lower share of left-handers was to be expected if the stigma of left-handedness and writing with the left hand has decreased over time. Although such a stigma was arguably no longer prevalent at the time of the hand grip measurement (2006 to 2014), the longer socialization of elderly individuals during times when prejudices were prevalent are likely to lead to the observed pattern.[10] Over the same period, the share of individuals who at least once report writing with their left hand increases first only mildly until the 1950 cohort, before rising sharply from then on, and eventually equaling the share of left-handers in the cohorts of the late 1990s.

Looking at the share of switched individuals, in the second graph of Figure 1.2, I find that about 90% of left-handers born before the end of World War Two write with their right hand today. This share rapidly decreased among those born in 1960 or later and reached zero by the late 1990's.

For at least 40 years (1920–1960), there existed a pooling equilibrium in which close to all left-handers were successfully switched and it can be reasonably assumed that a good fraction of non-switched left-handers at least tried to switch. Within less than another 40 years this practice vanished. Why did the norm of right-hand writing disappear so quickly? Young (1996) argues that previously stable conventions can suddenly reach a tipping point and change due to some

[10] The competing and somewhat prominent hypothesis that left-handers have a lower survival rate than right-handers (Coren and Halpern, 1991; Halpern and Coren, 1988) was quickly refuted, because the authors did not take stigma and switching among older cohorts into account (Harris, 1993).

Figure 1.2: Left-handedness and switching by year of birth



Note: Upper panel: Share of respondents who report innate left-handedness and write with their left hand by year of birth. Lower panel: Share of switched (if left-handed) by year of birth. Includes 95% confidence intervals.

idiosyncratic shock. Although only few agents are initially affected by the shock, a new generation of agents samples from the behavior of the previous one and adopts their convention according to the majority observed. Young's example is that of right-side driving in continental Europe, which was first introduced in France after the French Revolution and then spread by Napoleon into occupied territories, with neighboring countries successively following suit.

Here, such a shock may be the social upheavals of the 1960s, which found their climax in 1968 in Germany (*68er-Bewegung*) and elsewhere. The abandonment of conservative social norms, of which the correct writing hand is obviously just one, was initiated by this first generation of post-war raised parents.[11]

Another explanation may be technological change, for two reasons. First, labor became less manually intensive with a decline in the necessity to use tools and machinery which may be geared towards right-handers. Second, as production techniques became more flexible, goods whose functionality depends on manual handling (e.g., scissors and knives) were more and more produced for left-handers, making it unnecessary to switch hands.

1.4 Parental investment decisions

1.4.1 Theory

This section sets up a simple child investment model based on Becker and Tomes (1994). It incorporates social norms to deliver predictions about the relation between switching and parental characteristics.

Let d be a parental investment (switching). As in Becker and Tomes (1994), the parents' utility at time t depends on their own consumption of a good z_t minus

[11] A similar observation has been made by Coudé et al. (2006), who find a sharp increase in reported left-handedness among individuals who entered school shortly after the events of May 1968 in France. I do not find this increase in the share of reported left-handers, but instead a steep decline in switching rates.

the immediate treatment costs k plus their child's utility u_c. They try to maximize

$$u_t = u(z_t - kd) + \delta\, u_{ch}(y^d_{t+1}), \tag{1.1}$$

with respect to d, where δ is the parents' degree of altruism and y^d_{t+1} is some (discounted) future outcome of the child, such as its human capital. Consider only two points in time, t and $t+1$, and assume all utility functions are concave. I now incorporate the parents' social norms when making investment decisions, as proposed by Cunha and Heckman (2007, 2008). The parents may not know whether y^1_{t+1} or y^0_{t+1} is larger, but their expectations $\tilde{y}^d_{t+1}$ are influenced by a social norm $\bar{d}_t$ at t:

$$\tilde{y}^d_{t+1} = y(a_t) - c\,|\bar{d}_t - d|, \tag{1.2}$$

where $y\prime > 0$, c is a penalty term corresponding to the degree of conformity to the norm and a_t is a set of other important factors, such as parental education and resources, institutional policies, and innate ability. This formulation of conformity is borrowed from the model of social distance by Akerlof (1997). It reflects the idea that parents think that non-conformity with the existing norms leads to a possibly life-long penalty for the child due to stigmatization by teachers, peers and employers.[12] Note that I assume here that parents expect the norm to also exist at time $t+1$. The decision whether to switch the child's handedness is based on (1.1), where y^d_{t+1} is replaced with $\tilde{y}^d_{t+1}$. If $\bar{d}_t = 1$, then parents will switch their child if

$$u(z_t - k) + \delta\, u_c(y(a_t)) \geq u(z_t) + \delta\, u_c(y(a_t) - c) \tag{1.3}$$

$$\Leftrightarrow \delta\,(u_c(y(a_t)) - u_c(y(a_t) - c)) \geq u(z_t) - u(z_t - k). \tag{1.4}$$

[12] This is similar to Lindbeck and Nyberg (2006), who study the imposition of work norms by parents on their children.

Switching will thus be performed when the present forgone utility of doing it is smaller than the child's future gain, scaled by altruism. Obviously, no switching will take place if parents are selfish ($\delta = 0$), sardistic ($\delta < 0$), non-conformists ($c \leq 0$), or simply do not know that $\bar{d}_t = 1$.

What predictions can be derived from this simple model? As seen from (1.4), the threshold for switching the child is lower when parents have a high level of consumption, because then the difference $u(z_t) - u(z_t - k)$ is relatively lower for low levels of consumption. Altruism δ in general induces parents to invest in their child's future well-being. Hence, switching is merely one from among a range of measures that parents can undertake to foster their child's capabilities and standing in society. On the other hand, the difference $u_c(y(a_t)) - u_c(y(a_t) - c)$ decreases in endowments a_t. A stronger conformism to norms (expressed in a higher c) could lead to a negative empirical selection mechanism of switched individuals.

Lastly, the parent's switching decision may not be perfectly implementable. Instead, its probability of success depends on the child's already developed cognitive skills and motivation at time t. This concern is confirmed by Sattler (1996), who notes that it are usually the brighter and more motivated children on which switching attempts are successful. Empirically, this would link switching and skills in adulthood through a reverse causality. This step in the switching process is non-negligible. Using a world-wide survey of more than 11,000 individuals, Perelle and Ehrman (1994) find that switching attempts were successful in only 72% of the cases.

In summary, my parental investment decision model gives us some reasons to expect a positive selection of switched left-handers and some reasons to expect a negative. I now investigate these predictions empirically. Since I do not directly observe parental endowment, altruism, or conformism, I resort to parental education, degree of urbanization at age 15, and religious affiliation as proxies.

1.4.2 Empirics

Table 1.2 shows the results from a linear probability model regression of switching on control variables in the sample of left-handers. Starting with the basic demographic variables in column one, I find that females are 4.4 percentage points less likely to be switched. This difference is stable across cohorts, as shown in the upper graph of Figure 1.3.

East Germans are about 10 percentage points more likely to be forced to switch their writing hand. Anecdotal evidence suggests that in East Germany, as in other Eastern European ex-communist states, left-handers were suspected of being more creative than right-handers and hence as more likely to be a threat to the ruling regime. This threat was supposed to be eliminated by switching the writing hand Sattler (1996). I illustrate switching trends between West and East Germany in the lower graph of Figure 1.3. I find that the level difference between East and West is driven by cohorts born after 1960. The Berlin Wall was built in 1961 and brought a new wave of oppression. As dissidents could no longer simply leave the country, the regime aimed to stigmatize those who did not conform to the socialist ideology, starting early in school. Thus, the liberal movement of the 1960s was much less developed in the East than in the West (Ohse, 2010).

High paternal education negatively predicts switching, but the coefficients are not significant. In contrast, higher maternal education is significantly positively associated with switching. Highly educated mothers may have higher reputational concerns and like to see their children conform to existing norms.

Both coefficients of migration background are negative, but not significantly so. Column two of Table 1.2 excludes migrants from the sample and uses religious affiliation as an alternative explanatory variable. Individuals belonging to the Protestant Evangelical denomination are 7.6% less likely to be switched than Catholics. The dummy for 'other' affiliations (Islam and other Christian denominations) exhibits a large and positive coefficient which is not significant due to a low number of cases. The degree of urbanization at age 15 is never a significant predictor.

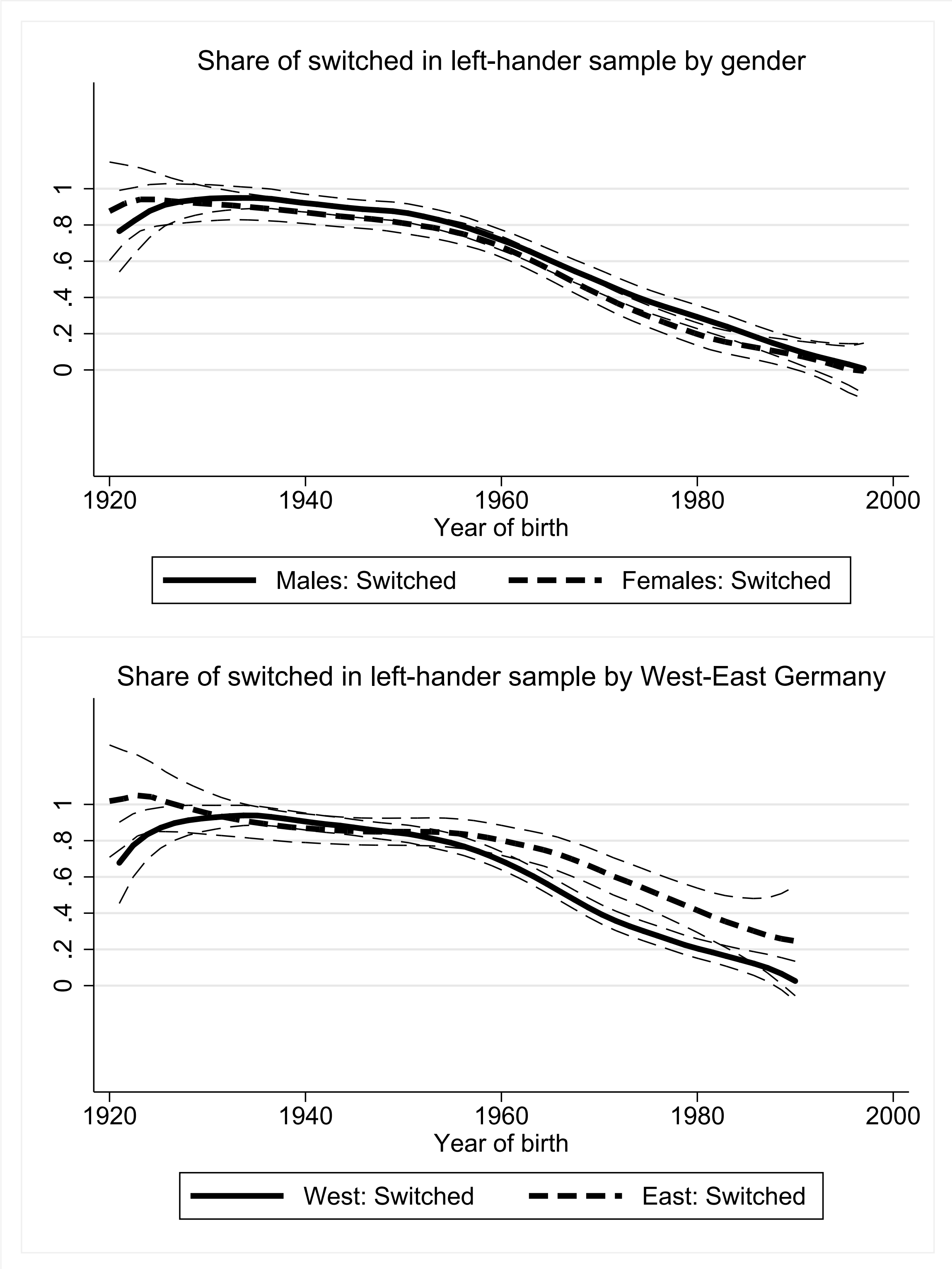
In conclusion, the regression analysis shows that gender, being East German, and maternal education are significant predictors of successful switching, confirming some predictions of the theoretical model.

Table 1.2: Regressions on switching indicator in left-hander sample

Sample restrictions:	(i)	(ii) Non-migrants only
Share of switched:	56.07%	58.14%
Female	-0.044*	-0.033
	(0.024)	(0.030)
East German	0.099***	0.075**
	(0.031)	(0.038)
Mothers education		
Basic/none	(ref.)	
Middle/other	0.077**	0.059
	(0.039)	(0.051)
High	0.103*	0.091
	(0.054)	(0.076)
Missing	-0.035	-0.133
	(0.066)	(0.094)
Fathers education		
Basic/none	(ref.)	
Middle/other	0.025	0.033
	(0.038)	(0.050)
High	-0.064	-0.075
	(0.049)	(0.063)
Missing	-0.017	0.026
	(0.056)	(0.084)
Urbanization at age 15		
Large city	(ref.)	
Mid-size city	-0.011	0.002
	(0.039)	(0.050)
Small town	-0.037	0.047
	(0.037)	(0.046)
Countryside	-0.001	0.027
	(0.035)	(0.043)
Missing	-0.061	0.128
	(0.104)	(0.176)
Migration background		
none	(ref.)	
1st Generation	-0.032	
	(0.037)	
2nd Generation	-0.039	
	(0.052)	
Religious affiliation		
Catholic		(ref.)
Protestant		-0.076**
		(0.035)
Denomination free		-0.011
		(0.046)
Other		0.158
		(0.111)
Missing		-0.077
		(0.047)
Cohort fixed effects	yes	yes
N	1,129	934
Adjusted R^2	0.399	0.401

Switched is an indicator equal to one if an individual self-reports at least once that he is born as a left-hander and writes with the right hand within one interview in any survey year. Robust standard errors in in parenthesis below. *** p<0.01, ** p<0.05, * p<0.1.

Figure 1.3: Switching by gender, East/West Germans and year of birth



Note: Upper panel: Share of switched (if left-handed) by gender and year of birth. Lower panel: Share of switched (if left-handed) by East/West Germans and year of birth. Includes 95% confidence intervals.

1.5 Results

In this section I investigate differences in later life (labor market) outcomes between left- and right-handers and between the three groups, switched left-handers, non-switched left-handers, and right-handers, where the last are the reference group in both cases. Though not the main focus of this paper, pure left–right differences are included for comparison with the existing literature.

I use OLS and random effects (RE) models for the estimation. Thus, my results do not allow for a causal interpretation, since the switching decision has to be made consciously by parents and teachers. It is thus a priori unlikely that the individual's characteristics that have been found (in the previous section) to (weakly) correlate with the intervention are its sole predictors. In particular, a possible reverse causality with respect to innate cognitive and non-cognitive skills is worrisome. Furthermore, and as derived by my theoretical model, parents with higher altruism and involvement with their child's development may contribute to a positive selection bias.

I undertake two measures to address these concerns. First, I show, in robustness checks, that my results do not change significantly when including control variables that are strong predictors of the outcome variables, in particular parental education. Second, in the next section I discuss a difference-in-difference strategy to identify the causal effect of switching, and find very similar results.

Nevertheless, the contrasts reported here are interesting enough on their own, because they can demonstrate associations of an endogenous childhood intervention with personality traits, cognitive skills, and economic performance in adulthood, something rarely observed. In particular, given that deficits in these quantities have been reported for left-handers in general, it is worthwhile to document that social norms can have any influence on innate and hard-wired skills. For now, the regression model takes the form

$$y_{it} = \alpha_L \, lefty_i + \beta_0 \, X_i + \mu_t + \epsilon_{it} \tag{1.5}$$

to investigate differences between left- and right-handers and

$$y_{it} = \alpha_N \, non\text{-}switched\ lefty_i + \alpha_S \, switched\ lefty_i + \beta_0 \, X_i + \nu_t + \delta_{it} \tag{1.6}$$

to investigate differences between switched and non-switched left-handers with respect to right-handers. The index t denotes the survey year and is irrelevant for all channel outcome variables which are cross-sectional. My preferred specification includes cohort fixed effects μ_t and all control variables X_i from column one of Table 1.2. Age fixed effects are included for labor market outcomes. All control variables are pre-determined before the intervention or non-changeable by the individual (such as parental education), to avoid bias from bad controls. In particular, I do not include occupation or industry controls. I use linear probability models for binary outcomes (employment status, higher track).

1.5.1 Labor market outcomes

Table 1.3 shows regressions on employment status and log-hourly wages. Starting with general left–right differences in the first panel, I find that left-handers are 2.3 percentage points less likely to be employed and if they are, they earn about 7% lower wages, which is close to the wage gap of 6% reported by Goodman (2014). The second panel in Table 1.3 splits left-handers into switched and non-switched. I find that it is the latter who perform significantly worse on both measures, being about 6 percentage points less likely to be employed, as well as earning 11% lower wages than right-handers. Switched left-handers earn a statistically insignificant 3% lower hourly wages.

Table 1.3: Baseline results for labor market outcomes

	(i)	(ii)
Outcome:	Employed	Log(Wage)
Pool left-handers:		
Left-handed	-0.023*	-0.069***
	(0.012)	(0.022)
Split up left-handers		
Switched lefty	0.005	-0.033
	(0.016)	(0.028)
Non-switched lefty	-0.057***	-0.115***
	(0.019)	(0.033)
Controls, cohort and age f.e.	yes	yes
N	53,213	43,514
N(cluster)	8,513	7,600
Overall R^2	0.061	0.164

Left-handed equals one if a respondent in the German SOEP reports at least once to be a natural left-hander during grip strength measurements performed in 2006, 2008, 2010, 2012 and 2014. Switched lefty equals one if the respondent reports at least once a difference between her natural and writing in the same year. The upper panel regresses outcome variables on controls and an indicator for being left-handed. The lower panel differentiates between switched and non-switched left-handers. In both cases, right-handers are the reference group. Table uses a linear random effects model for all outcomes. All regressions control for cohort fixed effects, whether the individual grew up in West or East Germany, gender, migration background (none, 1st generation, 2nd generation), mothers' and fathers' education (low/none, middle, high, missing), and urbanization at age 15 (large city, mid-size city, small town, countryside, missing). Sample restricted to individuals between age 25 and 60. Pools observations between years 2004 to 2014. Standard errors clustered at individual level in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 1.4: Channels: Schooling, full sample

	(i)	(ii)	(iii)	(iv)
Outcome:	Education	Higher track	Math grade	German grade
Unconditional mean	12.353	0.606	0.000	0.000
Pool left-handers:				
Left-handed	-0.069	-0.031**	-0.036	-0.029
	(0.083)	(0.015)	(0.039)	(0.039)
Adjusted R^2	0.251	0.252	0.033	0.113
Differentiate between switched and non-switched left-handers:				
Switched lefty	0.091	-0.004	0.047	-0.008
	(0.101)	(0.019)	(0.046)	(0.048)
Non-switched lefty	-0.339**	-0.076***	-0.159**	-0.062
	(0.141)	(0.025)	(0.067)	(0.063)
Controls	yes	yes	yes	yes
Cohort fixed effects	yes	yes	yes	yes
N	11,249	9,940	7,541	7,265
Adjusted R^2	0.252	0.252	0.034	0.113

The upper panel regresses outcome variables on controls and an indicator for being left-handed. The lower panel differentiates between switched and non-switched left-handers. In both cases, right-handers are the reference group. Table uses linear regression for all outcomes. All regressions control for cohort fixed effects, whether the individual grew up in West or East Germany, gender, migration background (none, 1st generation, 2nd generation), mothers' and fathers' education (low/none, middle, high, missing), and urbanization at age 15 (large city, mid-size city, small town, countryside, missing). Sample restricted to cohorts born between 1920 and 1997. Years of education includes only individuals of age greater or equal to 25 at time of observation in survey. Higher track is a dummy variable equaling one if completed schooling track is higher than the lowest (*Hauptschule*). Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 1.5: Channels: Schooling, wage sample

	(i)	(ii)	(iii)	(iv)
Outcome:	Education	Higher track	Math grade	German grade
Unconditional mean	12.799	0.723	-0.008	-0.045
Pool left-handers:				
Left-handed	-0.050	-0.023	-0.019	-0.001
	(0.105)	(0.018)	(0.050)	(0.049)
Adjusted R^2	0.227	0.159	0.0287	0.127
Differentiate between switched and non-switched left-handers:				
Switched lefty	0.205	0.009	0.110*	-0.009
	(0.135)	(0.024)	(0.061)	(0.065)
Non-switched lefty	-0.372**	-0.062**	-0.170**	0.007
	(0.160)	(0.027)	(0.077)	(0.071)
Controls	yes	yes	yes	yes
Cohort fixed effects	yes	yes	yes	yes
N	7,434	6,609	4,940	4,739
Adjusted R^2	0.228	0.159	0.030	0.127

See the notes of Table 1.3 for table description and the list of control variables, and Table 1.4 for the definition of the outcome variables. In contrast to the sample in Table 1.4, the sample in this table is restricted to individuals included in the regression on wages from column two of Table 1.3. Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

1.5.2 Channels

Human capital accumulation

I now investigate potential channels which could lead to the observed differences in long-run labor market outcomes. A natural start consists of measures of human capital, as writing is primarily learned in school and the "correct" writing hand might matter for school success. Table 1.4 shows the results, using the same model specification as for labor market outcomes (excluding age dummies).

Differences between left- and right-handers (upper panel) are rather small, and only significant for tracking in column two. Similar to labor market outcomes, the contrasts in human capital between switched and non-switched left-handers to right-handers are remarkably different. Non-switched lefties report about one-third fewer years of education, or 0.12 standard deviations (column one). Column two shows that they are also 7.6 percentage points (12.5%) less likely to graduate from a school track which is higher than the lowest. As hypothesized in Section 1.3, teachers might prohibit left-handers from pursuing higher school tracks when they fail to switch their writing hand to the right. Hence discrimination by teachers based on the signal of the writing hand could explain later life outcomes alone. However, non-switched left-handers also report significantly lower grades compared to right-handers and switched left-handers, which indicates that switching might be related to more than just conformity. Math grades are 0.16 standard deviations lower (column three) and German grades 0.06 standard deviations lower (column four), although the latter coefficient is insignificant.[13] It is unlikely that grades from the last school certificate are driven by discrimination by teachers, especially in Math. Grades in German show no significant differences (column four).

While Table 1.4 used the full sample, the results for the sample with positive wages are shown in Table 1.5. The coefficients are largely similar, but switched left-handers obtain 0.2 more years of education and a significant 0.1 standard

[13]The difference in Math grades is not driven by selection into different school tracks. Controlling for the latter leads to very similar results.

deviation higher Math grades than right-handers.

Cognitive and non-cognitive skills

Next, I investigate whether left-handedness and, in particular, switching, is also correlated with cognitive and non-cognitive skills in adulthood. Again I use the same regression specification as for human capital measures. The results are displayed in Table 1.6. Starting with my two measures of cognitive skills in columns one and two, I find that there are no significant differences either between left- and right-handers, nor when I split up left-handers into switched and non-switched.

To some extent, this comes as a surprise since the previous literature found that left-handers perform worse than right-handers in some of these tests. However, not in all of them. Johnston et al. (2009, 2013) find that left-handed children do not perform worse than right-handers in tests which require vocabulary and expressive language skills, which the animal naming test requires. Goodman (2014) reports that left-handers perform worse in reciting numbers backwards but not forwards, which requires short-term memory skills. Thus, the symbol-digit test might not capture the dimensions of cognitive skills in which left-handers perform worse. It is interesting to note, however, that neither the switching status is associated with a difference in these measures.

Roughly the same holds for the Big Five personality traits in columns three to seven of Table 1.6. Left-handers describe themselves as significantly less conscientious and extroverted, but are more neurotic. Less conscientiousness and a higher level of neuroticism could be correlated with the behavioral and learning problems in childhood which Goodman (2014) and Johnston et al. (2009, 2013) document for left-handers. Similar coefficients for switched and non-switched left-handers indicate that such behavioral problems arise for both types of left-handers. In contrast, there are significant differences in the external locus of control (LoC) (column eight). Switched left-handers have a significantly higher external LoC than right-handers. This is in line with Piatek and Pinger (2015), who find that the locus of control's influence on wages is mostly through education.

Table 1.7 repeats the previous regressions for employed individuals. As was the case for human capital variables, I find similar results, but in this sample, the contrasts between left- and right-handers are mainly driven by switched left-handers. They report significantly less conscientiousness, extraversion, agreeableness, and more neuroticism. This is in line with the psychological literature documenting that switched left-handers are often introverted and tend to disagree with others (Sattler, 1996).

Table 1.6: Channels: Cognitive and non-cognitive skills, full sample

	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)
	Cognitive skills		Non-cognitive skills					
	SDT	ANT	Openness	Conscient-iousness	Extra-version	Agreeable-ness	Neuroticism	External locus of control
Unconditional mean	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Pool left-handers:								
Left-handed	-0.017	-0.073	0.013	-0.068**	-0.072**	-0.026	0.078**	0.024
	(0.042)	(0.073)	(0.032)	(0.033)	(0.034)	(0.033)	(0.033)	(0.038)
Adjusted R^2	0.306	0.114	0.049	0.027	0.026	0.040	0.058	0.047
Differentiate between switched and non-switched left-handers:								
Switched lefty	-0.023	-0.087	0.015	-0.070*	-0.088**	-0.042	0.081*	-0.023
	(0.053)	(0.090)	(0.041)	(0.041)	(0.043)	(0.042)	(0.042)	(0.048)
Non-switched lefty	-0.006	-0.045	0.010	-0.065	-0.046	-0.000	0.072	0.104*
	(0.066)	(0.122)	(0.049)	(0.054)	(0.052)	(0.053)	(0.052)	(0.059)
Controls	yes	yes	yes	yes	yes	yes	yes	yes
Cohort fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
N	5,033	2,275	11,037	11,032	11,038	11,041	11,041	8,037
Adjusted R^2	0.306	0.113	0.049	0.027	0.026	0.040	0.058	0.048

See the notes of Table 1.3 for table description and the list of control variables. SDT refers to the sum of correct entries within 90 seconds in the symbol-digit test. ANT refers to the sum of uniquely named animals within 90 seconds in the animal naming test. Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism comprise the Big Five personality traits. See text for further details. Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 1.7: Channels: Cognitive and non-cognitive skills, wage sample

	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)
	Cognitive skills		Non-cognitive skills					
	SDT	ANT	Openness	Conscientiousness	Extraversion	Agreeableness	Neuroticism	External locus of control
Unconditional mean	0.238	0.186	0.056	0.092	0.027	-0.032	-0.038	-0.075
Pool left-handers:								
Left-handed	-0.052	-0.034	-0.004	-0.063*	-0.099**	-0.085**	0.083**	0.058
	(0.055)	(0.102)	(0.039)	(0.038)	(0.043)	(0.042)	(0.041)	(0.047)
Adjusted R^2	0.156	0.048	0.034	0.014	0.023	0.028	0.050	0.033
Differentiate between switched and non-switched left-handers:								
Switched lefty	-0.033	-0.088	-0.048	-0.092*	-0.136**	-0.134**	0.091*	0.010
	(0.071)	(0.127)	(0.052)	(0.049)	(0.057)	(0.055)	(0.053)	(0.062)
Non-switched lefty	-0.077	0.051	0.055	-0.024	-0.049	-0.018	0.072	0.130*
	(0.084)	(0.161)	(0.056)	(0.059)	(0.062)	(0.065)	(0.063)	(0.068)
Controls	yes	yes	yes	yes	yes	yes	yes	yes
Cohort fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
N	2,965	1,397	6,596	6,596	6,595	6,597	6,597	5,046
Adjusted R^2	0.156	0.048	0.035	0.014	0.023	0.029	0.050	0.033

See the notes of Table 1.3 for table description and the list of control variables. See the notes of Table 1.6 for the definition of the outcome variables. In contrast to the sample in Table 1.6, the sample in this table is restricted to individuals included in the regression on wages from column two of Table 1.3. Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

1.5.3 Channel analysis

I now set out to understand which of the channels considered in the previous section are responsible for the observed differences in labor market outcomes. To do this, I step-by-step include the channel variables which showed a significantly difference for either switched or non-switched left-handers to right-handers.[14] Thus, the regression model becomes

$$y_{it} = \alpha_L \, lefty_i + \beta \, X_i + \tau \, M_i + \mu_t + \epsilon_{it} \tag{1.7}$$

to investigate how the differences between left- and right-handers, but becomes

$$y_{it} = \alpha_N \, non\text{-}switched \; lefty_i + \alpha_S \, switched \; lefty_i + \beta \, X_i + \tau \, M_i + \nu_t + \delta_{it} \tag{1.8}$$

to investigate how the differences between switched and non-switched left-handers and right-handers change after inclusion of channel, or mediating, variables M.

To interpret the coefficients τ as causal would require the sequential ignorability assumption to hold (Imai et al., 2010): Conditional on values of some treatment T and pre-treatment covariables X, the mediating variables M have to be independent of potential outcomes $Y(t,m)$.[15] Given that the mediating variables appeared as outcome variables in the previous section, this assumption is unlikely to hold here. Therefore, I do not attempt to conduct a fully elaborate mediation analysis, but provide a mere descriptive analysis on the role of schooling and skills for the observed differences in labor market performance by left-handedness and switching status.

Note that not all channel variables are observed for all individuals from Table 1.3. As a consequence, the sample size is reduced by about 20%, which raises

[14] I do not include Math grades as a channel, as the number of observations for this outcome is very low.

[15] For completeness, the sequential ignorability assumption by Imai et al. (2010) also requires that T is independent of potential outcomes given X, which corresponds to a selection-on-observables assumption.

worries about sample selection. However, I show that my baseline results still hold in the most restricted sample and in samples for which only one group of channel variables is observed.[16]

I start with log-hourly wages. Column one in Table 1.8 repeats column two of Table 1.3 in the sample where all mediator variables are observed. Similarly to the results in the full sample, I find that non-switched left-handers earn 10% lower wages than right-handers, while the difference for switched left-handers is an insignificant -5%. Including years of education as mediator in column two leads both coefficients to become equal in size, -6%. Including only the Big Five personality traits (column three) instead of education leads to rather little change in the wage gaps, and so does adding only the external locus of control (column four). Inclusion of either set of channels reduces the wage gap of non-switched left-handers to 8%. Including all mediators in column five leads to the same result, qualitatively, as in column two: both switched and non-switched left-handers experience nearly the same wage gap of 5% to 6% to right-handers. The overall difference between all left- and right-handers (upper panel) of -6% changes remarkably little throughout this exercise, while the overall R-squared increases from 0.171 to 0.281.

To check whether these results are driven by sample selection, I repeat the analysis for samples where either only years of education or only personality traits, or both, are non-missing. The external locus of control is not considered because it leads to the largest loss in sample size. Columns one and two in Table 1.9 replicate the same columns from Table 1.8 for the larger sample and I find that their main finding still holds: taking into account human capital differences in column two, the wage gaps for switched and non-switched left-handers become much more similar. The sample in columns three and four of Table 1.9 is restricted to observations with non-missing values on personality traits. As in column four of Table 1.8, I find that

[16] Furthermore, in an unreported logit regression, I find that, conditional on being in the baseline samples of Table 1.3, the probability of being included in the corresponding restricted sample does not differ significantly between right-handers, switched and non-switched left-handers after controlling for cohort effects.

the wage gaps do not react to the inclusion of these traits. Note however, that also R-squared increases only marginally, from 0.161 to 0.167. In a sample where years of education and personality traits (columns five to seven) are both non-missing, I again observe that human capital differences are more important in reducing the difference in the wage gap than are personality traits. The results for employment status are provided in Table 1.10. Comparable to the baseline results in Table 1.3, I find that non-switched left-handers are 4% less likely to be employed than right-handers. This gap becomes an insignificant 3% after controlling for years of education (columns two and five). Again, non-cognitive skills do not matter as much as human capital. However, and in contrast to wages, switched left-handers are never significantly more or less likely to be employed than right-handers, even after taking into account human capital, personality, and the external locus of control.

Table 1.8: Log(wage) including channels (Random Effects model)

	(i)	(ii)	(iii)	(iv)	(v)
Outcome:			Log(Wage)		
Pool left-handers:					
Left-handed	-0.065**	-0.063***	-0.062**	-0.059**	-0.060**
	(0.026)	(0.024)	(0.026)	(0.025)	(0.024)
Overall R^2	0.171	0.262	0.177	0.203	0.281
Differentiate between switched and non-switched left-handers:					
Switched lefty	-0.046	-0.063**	-0.044	-0.045	-0.064**
	(0.033)	(0.031)	(0.033)	(0.033)	(0.031)
Non-switched lefty	-0.096**	-0.063*	-0.091**	-0.080**	-0.053
	(0.039)	(0.036)	(0.038)	(0.038)	(0.035)
Years of education		0.080***			0.075***
		(0.003)			(0.003)
Openness			0.020**		-0.012
			(0.009)		(0.008)
Conscientiousness			0.007		0.017**
			(0.009)		(0.008)
Extraversion			-0.008		0.006
			(0.009)		(0.008)
Agreeableness			-0.018**		-0.029***
			(0.008)		(0.007)
Neuroticism			-0.051***		-0.013*
			(0.008)		(0.007)
External locus of control				-0.125***	-0.090***
				(0.008)	(0.007)
Controls	yes	yes	yes	yes	yes
Cohort fixed effects	yes	yes	yes	yes	yes
Age fixed effects	yes	yes	yes	yes	yes
N	34,836	34,836	34,836	34,836	34,836
N(cluster)	4,918	4,918	4,918	4,918	4,918
Overall R^2	0.171	0.262	0.177	0.203	0.281

See the notes of Table 1.3 for table description and the list of control variables. Coefficients of channel variables are not shown in the upper panel. Table uses a linear random effects model. Sample restricted to individuals between age 25 and 60, with positive wages, and with non-missing values for years of education, Big Five personality traits, and external locus of control. Pools observations between years 2004 to 2014. See text for further details. Standard errors clustered at individual level in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 1.9: Log(wage) including channels years of education and personality (Random Effects model)

	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)
Outcome:				Log(Wage)			
Sample	Education only		Personality only		Education + Personality		
Pool left-handers:							
Left-handed	-0.067***	-0.064***	-0.060***	-0.057**	-0.059**	-0.058***	-0.054**
	(0.022)	(0.021)	(0.023)	(0.023)	(0.024)	(0.022)	(0.022)
Overall R^2	0.164	0.255	0.161	0.167	0.161	0.252	0.256
Differentiate between switched and non-switched left-handers:							
Switched lefty	-0.037	-0.054**	-0.027	-0.025	-0.032	-0.049*	-0.046*
	(0.029)	(0.027)	(0.030)	(0.030)	(0.030)	(0.028)	(0.028)
Non-switched lefty	-0.106***	-0.077**	-0.105***	-0.102***	-0.099***	-0.071**	-0.066**
	(0.034)	(0.031)	(0.036)	(0.035)	(0.037)	(0.033)	(0.033)
Years of education		0.079***				0.079***	0.079***
		(0.002)				(0.003)	(0.003)
Openness				0.014*			-0.018**
				(0.008)			(0.007)
Conscientiousness				0.007			0.022***
				(0.008)			(0.007)
Extraversion				-0.008			0.010
				(0.008)			(0.007)
Agreeableness				-0.019***			-0.025***
				(0.007)			(0.007)
Neuroticism				-0.054***			-0.037***
				(0.007)			(0.007)
Controls	yes	yes	yes	yes	yes	yes	yes
Cohort fixed effects	yes	yes	yes	yes	yes	yes	yes
Age fixed effects	yes	yes	yes	yes	yes	yes	yes
N	42,974	42,974	40,042	40,042	39,548	39,548	39,548
N(cluster)	7,434	7,434	6,595	6,595	6,448	6,448	6,448
Overall R^2	0.164	0.255	0.161	0.167	0.161	0.252	0.256

See the notes of Table 1.3 for table description and the list of control variables. Sample restricted to individuals between age 25 and 60, and with positive wages. Pools observations between years 2004 to 2014. Coefficients of channel variables are not shown in the upper panel. In columns one and two the sample is restricted to observations with non-missing years of education. In columns three and four the sample is restricted to observations with non-missing personality traits. The sample in columns five to seven is restricted to observations with non-missing values of both, years of education and personality traits. Table uses random effects regression. Standard errors clustered at individual level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.10: Employment status including channels (Random Effects model)

	(i)	(ii)	(iii)	(iv)	(v)
Outcome:			Employed		
Pool left-handers:					
Left-handed	-0.005	-0.003	-0.001	-0.003	-0.000
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Overall R^2	0.059	0.073	0.073	0.071	0.093
Differentiate between switched and non-switched left-handers:					
Switched lefty	0.020	0.016	0.022	0.019	0.017
	(0.018)	(0.017)	(0.018)	(0.018)	(0.017)
Non-switched lefty	-0.043*	-0.034	-0.037*	-0.038*	-0.027
	(0.023)	(0.022)	(0.022)	(0.023)	(0.022)
Years of education		0.019***			0.017***
		(0.002)			(0.002)
Openness			0.000		-0.007
			(0.005)		(0.005)
Conscientiousness			0.040***		0.039***
			(0.005)		(0.005)
Extraversion			0.006		0.008*
			(0.005)		(0.005)
Agreeableness			-0.013***		-0.015***
			(0.004)		(0.004)
Neuroticism			-0.023***		-0.012***
			(0.004)		(0.005)
External locus of control				-0.044***	-0.030***
				(0.004)	(0.005)
Controls	yes	yes	yes	yes	yes
Cohort fixed effects	yes	yes	yes	yes	yes
Age fixed effects	yes	yes	yes	yes	yes
N	41,876	41,876	41,876	41,876	41,876
N(cluster)	5,430	5,430	5,430	5,430	5,430
Overall R^2	0.060	0.073	0.074	0.071	0.093

See the notes of Table 1.3 for table description and the list of control variables. Coefficients of channel variables are not shown in the upper panel. Table uses random effects regression. Sample restricted to individuals between age 25 and 60, and with non-missing values for years of education, Big Five personality traits, and external locus of control. Pools observations between years 2004 to 2014 Standard errors clustered at individual level in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

1.5.4 Robustness checks

Robustness checks with respect to the set of control variables and samples for wages are provided in Table 1.11. Without any controls (column one), there are no significant differences between left- and right-handers. This is due to the variation of left-handedness in certain demographic groups, as discussed in Appendix 1.B. Controlling for these basic characteristics (gender, migration status, East German) in column two, I find a wage gap of 6%. Further controlling for the remaining co-variables leads to a slight increase to 7% in column five, which is equal to column two of Table 1.3. This finding is reassuring, as it demonstrates the randomness of left-handedness within certain demographic groups. In particular, including cohort fixed effects and parental education changes the coefficient only marginally while raising the overall R-squared substantially, from 9.7% (column two) to 14.7% (column four). To address the issue of endogenous reporting behavior among individuals born before 1950, I restrict the sample to observations born in 1950 or later, but include all control variables (column five). The wage gaps remain unchanged. Excluding individuals with some migration background, who are less likely to report left-handedness than natives, leads to an even larger wage gap of roughly 8% (column six). On the other hand, excluding East Germans, who are also less likely to report left-handedness, leaves the baseline results unchanged (column seven).

Throughout all specifications and in all sub-samples, non-switched left-handers earn significantly lower wages than right-handers, while switched left-handers do not earn significantly less. As expected from my analysis in Section 1.4, columns two to four show that basic demographic characteristics and cohort effects are the most important confounders for the switching status. Parental background again matters little. In the sample of natives (column six), the wage deficit for non-switched left-handers is even larger, at 14%. The same robustness checks for employment status are provided in Table 1.12. I find that the coefficients are very stable throughout this exercise, even more than they are for wages. However, the control variables are also less powerful in explaining variation in employment status than in wages.

Since human capital was found to be the main mediating factor between switching status and wages, I also conducted robustness checks for years of education in Table 1.13. After including basic demographics, all coefficients become stable. It is noteworthy that including parental education nearly quadruples the adjusted R-squared, from 6.4% to 24.8%, but the coefficient for non-switched left-handers remains virtually unchanged (columns three to four). When excluding individuals born before 1950, I find a significantly positive difference of switched left-handers to right-handers (column five). One explanation could be that the "missing" left-handers in the older cohorts are actually switched left-handers who report right-handedness.

Finally, I investigate the sensitivity of my results with respect to the definition of being left-handed and switched. My preferred measure categorizes individuals as being left-handed (switched) if they ever report being a natural left-hander (ever report a difference between the innate hand and the writing hand) in any survey wave. As alternatives, I can assign an individual to the left-handed (switched) group if she reports left-handedness (a difference between innate and writing hand) at least half the time or every time when asked about her natural handedness and writing hand. I thus created two additional indicators, left-handed (50%) for the former and left-handed (100%) for the latter case, and analogous indicators for switching. Under the first (second) alternative definition, 7.6% (5.6%) of respondents in the full sample are left-handed, of which 52.2% (44.2%) are switched. Table 1.14 contains these robustness checks. Compared with my baseline results from Table 1.3, the alternative definitions of my key variables make virtually no difference for labor market outcomes. The coefficients for employment status (columns one and two) are virtually unchanged and wage gaps between left- and right-handers and between non-switched left-handers and right-handers are even more pronounced (columns three and four). However, differences in human capital (columns five to eight) are lower than under my preferred assignment rule and less statistically significant. One possible explanation for this finding is positive selection bias. Left-handers, in particular those who were not switched, are more likely to report left-handedness every time they are asked if they had less experience of discrimination, e.g., in school which would also lead to lower years of education. Thus, this finding corroborates my interpretation of discrimination in school based on the writing hand.

Table 1.11: Robustness checks for log-hourly wages (Random Effects model)

	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)
Outcome:	Log(Wage)						
Sample	Full				Born $\geq$ 1950	Non-migrants	West-Germans
Pool left-handers:							
Left-handed	-0.017	-0.059***	-0.064***	-0.065***	-0.068***	-0.078***	-0.068***
	(0.024)	(0.023)	(0.022)	(0.022)	(0.022)	(0.024)	(0.024)
Overall R^2	0.000	0.096	0.122	0.147	0.161	0.173	0.154
Differentiate between switched and non-switched left-handers:							
Switched lefty	0.041	0.017	-0.030	-0.031	-0.027	-0.030	-0.031
	(0.032)	(0.029)	(0.029)	(0.029)	(0.029)	(0.030)	(0.032)
Non-switched lefty	-0.088**	-0.153***	-0.105***	-0.108***	-0.114***	-0.142***	-0.105***
	(0.035)	(0.034)	(0.034)	(0.033)	(0.033)	(0.038)	(0.034)
Controls:							
Demographics	no	yes	yes	yes	yes	yes	yes
Cohort fixed effects	no	no	yes	yes	yes	yes	yes
Parental education	no	no	no	yes	yes	yes	yes
Age fixed effects	no	no	no	no	yes	yes	yes
Urbanization at 15	no	no	no	no	yes	yes	yes
N	43,514	43,514	43,514	43,514	41,894	35,291	32,971
N(cluster)	7,600	7,600	7,600	7,600	7,104	6,003	5,857
Overall R^2	0.000	0.097	0.122	0.147	0.161	0.174	0.155

Robustness check to regression on log-wages in Table 1.3. See the notes of Table 1.3 for table description and the list of control variables. Table uses a linear random effects model. Sample restricted to individuals between age 25 and 60. Pools observations between years 2004 to 2014. Demographic controls are gender, migration background (none, 1st generation, 2nd generation) and an East German dummy. Standard errors clustered at individual level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.12: Robustness checks for employment status (Random Effects model)

	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)
Outcome:	Employed						
Sample	Full				Born $\geq$ 1950	Non-migrants	West-Germans
Pool left-handers:							
Left-handed	-0.009	-0.022*	-0.024*	-0.024*	-0.029**	-0.025*	-0.021
	(0.013)	(0.013)	(0.012)	(0.012)	(0.013)	(0.014)	(0.014)
Overall R^2	0.000	0.029	0.053	0.058	0.056	0.050	0.073
Differentiate between switched and non-switched left-handers:							
Switched lefty	0.024	0.008	0.004	0.003	-0.004	0.001	-0.003
	(0.016)	(0.016)	(0.016)	(0.016)	(0.017)	(0.018)	(0.019)
Non-switched lefty	-0.048**	-0.058***	-0.058***	-0.056***	-0.056***	-0.060***	-0.040**
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.021)	(0.019)
Controls:							
Demographics	no	yes	yes	yes	yes	yes	yes
Cohort fixed effects	no	no	yes	yes	yes	yes	yes
Parental education	no	no	no	yes	yes	yes	yes
Age fixed effects	no	no	no	no	yes	yes	yes
Urbanization at 15	no	no	no	no	yes	yes	yes
N	53,213	53,213	53,213	53,213	50,885	42,335	40,390
N(cluster)	8,513	8,513	8,513	8,513	7,838	6,648	6,574
Overall R^2	0.001	0.030	0.053	0.059	0.056	0.051	0.074

Robustness check to regression on employment in Table 1.3. See the notes of Table 1.3 for table description and the list of control variables. Table uses random effects regression. Sample restricted to individuals between age 25 and 60. Pools observations between years 2004 to 2014. Demographic controls are gender, migration background (none, 1st generation, 2nd generation) and an East German dummy. Standard errors clustered at individual level in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 1.13: Robustness checks for years of education (OLS)

	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)
Outcome:	Years of education						
Sample	Full				Born $\geq$ 1950	Non-migrants	West-Germans
Pool left-handers:							
Left-handed	0.036	-0.035	-0.068	-0.062	-0.053	-0.138	-0.071
	(0.096)	(0.095)	(0.093)	(0.084)	(0.101)	(0.092)	(0.094)
Adjusted R^2	0.000	0.022	0.064	0.248	0.234	0.238	0.281
Differentiate between switched and non-switched left-handers:							
Switched lefty	0.075	-0.014	0.116	0.094	0.234*	0.009	0.126
	(0.117)	(0.116)	(0.114)	(0.102)	(0.134)	(0.108)	(0.116)
Non-switched lefty	-0.029	-0.069	-0.379**	-0.326**	-0.369**	-0.411**	-0.355**
	(0.157)	(0.155)	(0.155)	(0.140)	(0.148)	(0.162)	(0.152)
Controls:							
Demographics	no	yes	yes	yes	yes	yes	yes
Cohort fixed effects	no	no	yes	yes	yes	yes	yes
Parental education	no	no	no	yes	yes	yes	yes
Urbanization at 15	no	no	no	no	yes	yes	yes
N	11,249	11,249	11,249	11,249	7,707	9,001	8,635
Adjusted R^2	0.000	0.025	0.064	0.248	0.235	0.238	0.282

Robustness check to regressions in Table 1.4. See the notes of Table 1.3 for table description and the list of control variables. Demographic controls are gender, migration background (none, 1st generation, 2nd generation) and an East German dummy. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 1.14: Robustness checks by definition of left-handedness and switching

Outcome:	(i) Employed (RE)	(ii)	(iii) Log-wage (RE)	(iv)	(v) Education (OLS)	(vi)	(vii) Higher track (OLS)	(viii)
Pool left-handers:								
Left-handed(50%)	-0.024*		-0.080***		-0.075		-0.026	
	(0.013)		(0.024)		(0.090)		(0.016)	
Left-handed(100%)		-0.027*		-0.071**		-0.040		-0.027
		(0.015)		(0.028)		(0.105)		(0.019)
Overall/Adjusted R^2	0.061	0.061	0.164	0.163	0.251	0.251	0.252	0.252
Differentiate between switched and non-switched left-handers:								
Switched lefty(50%)	0.002		-0.028		0.037		-0.005	
	(0.018)		(0.032)		(0.110)		(0.021)	
Non-switched lefty(50%)	-0.050***		-0.133***		-0.238		-0.055**	
	(0.019)		(0.035)		(0.150)		(0.025)	
Switched lefty(100%)		-0.006		-0.004		0.035		-0.002
		(0.023)		(0.040)		(0.133)		(0.025)
Non-switched lefty(100%)		-0.046**		-0.120***		-0.095		-0.044
		(0.021)		(0.038)		(0.162)		(0.027)
Controls	yes	yes	yes	yes	yes	yes	yes	yes
Cohort fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
N	53,213	53,213	43,514	43,514	11,249	11,249	9,940	9,940
N(cluster)	8,513	8,513	7,600	7,600	11,249	11,249	9,940	9,940
Overall/Adjusted R^2	0.061	0.061	0.164	0.164	0.251	0.251	0.252	0.252

Robustness checks with respect to the definition of being left-handed and switched. See the notes of Table 1.3 for table description and the list of control variables. Left-handed(Switched) (50%) indicates individuals that self-report to be naturally left-handed (report a difference between their innate and writing hand) at least half the time across survey waves. Left-handed (Switched) (100%) indicates individuals that always self-report to be naturally left-handed (always report a difference between their innate and writing hand) across all survey waves. Table uses linear random effects models in columns (i) to (iv) and OLS in columns (v) to (viii). Sample restricted to individuals between age 25 and 60. Pools observations between years 2004 to 2014. Clustered standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

1.6 Discussion

The analysis so far has been limited to comparing conditional means between endogenous groups defined by handedness and switching status; this precludes claims about causality. As discussed above, one worry is positive selection bias of switched left-handers due to unobserved parental characteristics, and another is reverse causality from endowed skills which increase the likelihood of a successful switching. As motivated from my theoretical model of parental investment in Section 1.4, a variation in the prevalence of the switching practice across cohorts can aid in developing an identification strategy.

Since right-handers were never switched, their cohort trend can serve as a counterfactual for left-handers. The identifying assumption in this approach is that time trends for left- and right-handers would have developed in parallel in the absence of switching. I am not aware of any institutional change in schooling practices which applied only to left-handers. However, if trends would have changed for other reasons, I would falsely attribute this change to switching. For example, stigmatization and prejudices against left-handers in society may decline in general, leading left-handers to catch up over time. Decline of the pathological left-hander due to improvements in perinatal medical care over time could also be a confounding trend.

The prevalence of switching might thus only serve as an indicator for a positive development of attitudes towards left-handedness and liberal schooling practices. Violation of the exclusion restriction would lead to a downward bias of the IV estimate.[17] Since the OLS estimate is likely to be upward biased (due to positive selection), the OLS and IV estimate could serve as upper and respectively lower bounds on the true causal effect (Nevo and Rosen, 2012).

Under the assumption that switching is the only reason for differing cohort

[17] To see this, note that the reduced form coefficient, the difference in cohort trends between left- and righthanders, will be positive if e.g. stigmatization in school decreases over time. The first stage coefficient is negative as switching decreases over time. Thus, the IV estimate will be downward biased.

trends for left- and right-handers, I can employ a two-stage least-squares (2SLS) estimation, with a difference-in-differences specification in the first stage. I formalize this idea in the following two-equation model

$$Y_{it} = \alpha\, switching_i + \gamma_1 lefty_i + \theta_1\, \phi(t_i) + \beta_1\, X_i + \epsilon_{it} \tag{1.9}$$

$$switching_{it} = \delta\, \phi(t_i) \times lefty_i + \gamma_0 lefty_i + \theta_0\, \phi(t_i) + \beta_0\, X_{it} + \nu_{it}, \tag{1.10}$$

where α is the treatment effect of interest, the X_i are control variables, and $\phi(t_i)$ is a function of the cohort trend, e.g., a linear or quadratic trend. The dummy $switching_i$ denotes switched individuals, $lefty_i$ denotes reported left-handedness, and ϵ_{it} and ν_{it} are two correlated error terms. The set of control variables X_i is the same as in the previous section. In what follows I use a quadratic cohort trend, hence $\phi(t_i) = \nu_1\, t_i + \nu_2\, {t_i}^2$.

Table 1.15 shows estimates of α and γ_1 using OLS and 2SLS regressions for labor market outcomes.[18] I find that the OLS and 2SLS coefficients are very similar for both outcome variables. While the standard errors are higher from the 2SLS estimation, these results suggest only minor problems of selection bias for labor market outcomes. The first stage F-statistics are quite large, with a value of 120.

Tables 1.16 and 1.17 contain robustness checks for the 2SLS estimation similar to those in section 1.5, for employment and wages respectively. Across all specifications, I again find that the effect of left-handedness is negative on both outcomes, while the effect of switching is positive. For employment, the effect size of switching is similar to that of left-handedness, while it is slightly smaller for log-wages as in table 1.15. However, despite large first stage F-statistics, the coefficient for switching is imprecisely estimated and never significant at conventional levels.

[18] Note that the difference in the OLS coefficients between Tables 1.3 and 1.15 comes from the fact that all switched individuals are also left-handed. It follows that the coefficient on left-handedness in Table 1.15 corresponds to the non-switched left-hander coefficient in Table 1.3.

Table 1.15: OLS and 2SLS estimates for labor market outcomes

	(i)	(ii)	(iii)	(iv)
	Employed		Log(Wage)	
	OLS	2SLS	OLS	2SLS
Unconditional mean	0.820	0.820	2.596	2.596
Left-handed	-0.049***	-0.050*	-0.096***	-0.101**
	(0.019)	(0.029)	(0.032)	(0.048)
Switched	0.054**	0.055	0.063	0.072
	(0.024)	(0.052)	(0.043)	(0.084)
Controls	yes	yes	yes	yes
Cohort fixed effects	yes	yes	yes	yes
Age fixed effects	yes	yes	yes	yes
N	53,213	53,213	43,514	43,514
N(cluster)	8,513	8,513	7,600	7,600
Adjusted R^2	0.061	0.061	0.164	0.164
First stage F-stat		121.33		119.91

See table 1.3 for control variables and variable definition. The dummy variable switched is instrumented by the interactions of an indicator for being left-handed and a linear and quadratic cohort trend (two instruments). The first stage is visualized in the lower graph of Figure 1.2. Right-handers were never switched. In effect, cohort trends of right-handers serve as counterfactual. Standard errors clustered at individual level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.16: Robustness checks for employment status in 2SLS estimation

	(i)	(ii)	(iii)	(iv)
Outcome:	Employed			
Sample	Full	Born $\geq$ 1950	Non-migrants	West-Germans
Left-handed	-0.050*	-0.044	-0.054	-0.032
	(0.030)	(0.031)	(0.034)	(0.029)
Switched	0.054	0.041	0.046	0.030
	(0.053)	(0.056)	(0.059)	(0.055)
Controls:				
Demographics	yes	yes	yes	yes
Cohort fixed effects	yes	yes	yes	yes
Parental education	no	yes	yes	yes
Urbanization at 15	no	yes	yes	yes
N	53,213	50,885	42,335	40,390
N(cluster)	8,513	7,838	6,648	6,574
Adjusted R^2	0.053	0.056	0.050	0.073
First stage F-stat	121.04	109.92	97.97	111.09

Robustness checks for 2SLS regression on employment from Table 1.15. Demographic controls are gender, migration background and an East German dummy. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.17: Robustness checks for log-hourly wages in 2SLS estimation

	(i)	(ii)	(iii)	(iv)
Outcome:	Log(Wage)			
Sample	Full	Born $\geq$ 1950	Non-migrants	West-Germans
Left-handed	-0.085*	-0.105**	-0.122**	-0.103**
	(0.049)	(0.050)	(0.057)	(0.050)
Switched	0.050	0.082	0.089	0.088
	(0.086)	(0.089)	(0.093)	(0.091)
Controls:				
Demographics	yes	yes	yes	yes
Cohort fixed effects	yes	yes	yes	yes
Parental education	no	yes	yes	yes
Urbanization at 15	no	yes	yes	yes
N	43,514	41,894	35,291	32,971
N(cluster)	7,600	7,104	6,003	5,857
Adjusted R^2	0.122	0.161	0.173	0.154
First stage F-stat	119.49	109.95	94.58	105.23

Robustness checks for 2SLS regression on log-wages from Table 1.15. Demographic controls are gender, migration background and an East German dummy. *** p<0.01, ** p<0.05, * p<0.1.

1.7 Conclusion

Recent research documents that left-handers have lower cognitive skills than right-handers, are more likely to have behavioral and learning problems as children, and perform worse on the labor market. These differences are likely to be of a pathological origin (Goodman, 2014). Does an early childhood intervention such as switching compensate or engrave such deficits?

I find evidence for the former hypothesis. In contrast to non-switched left-handers, those left-handers that write with their right hand today do not show lower measures of human capital than right-handers. Labor market outcomes show a similar pattern: non-switched left-handers earn about 10% to 11% lower wages than right-handers and are less likely to be employed. There exist no statistically significant differences between those forced to switch and right-handers.

One explanation for this observation could be that switching the writing hand is a signal of conformism and ability, one that can be observed by teachers who would prohibit students' progression to higher tracking schools in the absence of this signal. Another explanation could be that these children receive additional attention and care by their teachers and parents. A successful switching of the writing hand may also induce a feeling of success and motivate children to improve their skills further in the future.

My findings are robust to applying different sets of control variables, excluding specific sub-samples and to the definition of left-handedness and switching. Furthermore, I use the cohort trends of right-handers as a counter-factual for the cohort trends of left-handers in a difference-in-difference approach. The interaction of left-handedness and cohort trends delivers a strong first stage as switching declines rapidly across time and was never performed on right-handers. My identification strategy requires that the cohort trends of labor market variables would have evolved in parallel between left- and right-handers. I argue that if they did not, IV estimates will deliver a lower bound on the true effect, while OLS estimates deliver an upper bound due to positive selection into switching. However, I find that the IV estimates are close to those using OLS, suggesting that selection bias is of minor concern.

I also document differences in the Big Five personality traits and external locus of control by handedness and switching status. However, these are not primarily responsible for the observed wage gaps. No significant differences were found for two measures of cognitive skills, which somewhat stands in contrast to the existing literature that reported left-right-hander differences among children (Johnston et al., 2009, 2013; Goodman, 2014). Including the intermediate variables step-by-step into the log-wage regression reveals that labor market gaps are driven by human capital accumulation.

My findings point to the importance and long run persistence of early infant endowments and the compensatory function of early childhood interventions in a vulnerable population.

Appendix

1.A Personality and the Locus of Control

Table 1A.1 shows the wording of the items that were used to construct the Big Five personality traits and the external locus of control. Respondents could answer on a Likert-type scale ranging from 1 (does not apply at all) to 7 (applies completely). The corresponding answers were averaged and standardized in the analysis sample.

Table 1A.1: SOEP items used to construct Big Five personality traits and Locus of Control

Item label	Trait
Big Five: I see myself as someone who...	
is original, comes up with new ideas	Openness
values artistic experiences	Openness
has an active imagination	Opennes
does a thorough job	Conscientiousness
does things effectively and efficiently	Conscientiousness
is rather lazy (reversed) [not used]	Conscientiousness
is communicative, talkative	Extraversion
is outgoing, sociable	Extraversion
is reserved (reversed) [not used]	Extraversion
is sometimes somewhat rude to others (reversed)	Agreeableness
has a forgiving nature	Agreeableness
is considerate and kind to others	Agreeableness
worries a lot	Neuroticism
gets nervous easily	Neuroticism
is relaxed, handles stress well (reversed)	Neuroticism
Locus of control	
Compared to other people, I have not achieved what I deserve.	External LOC
What a person achieves in life is above all a question of fate or luck.	External LOC
I frequently have the experience that other people have a controlling influence over my life.	External LOC
The opportunities that I have in life are determined by the social conditions.	External LOC
Inborn abilities are more important than any efforts one can make.	External LOC
I have little control over the things that happen in my life. [not used]	External LOC

Note: Table follows Heineck and Anger (2010)

1.B Reporting left-handedness

Although left-handedness is nearly random in the population (Johnston et al., 2009), self-reporting it is not. As mentioned in Section 1.3, the previous literature has used more and more sensitive measures to construct the variables for handedness than I are able to use. This section thus sets out to investigate which characteristics predict reported left-handedness and are thus important to avoid bias from measurement error. One additional predictor that I use is country of birth[19].

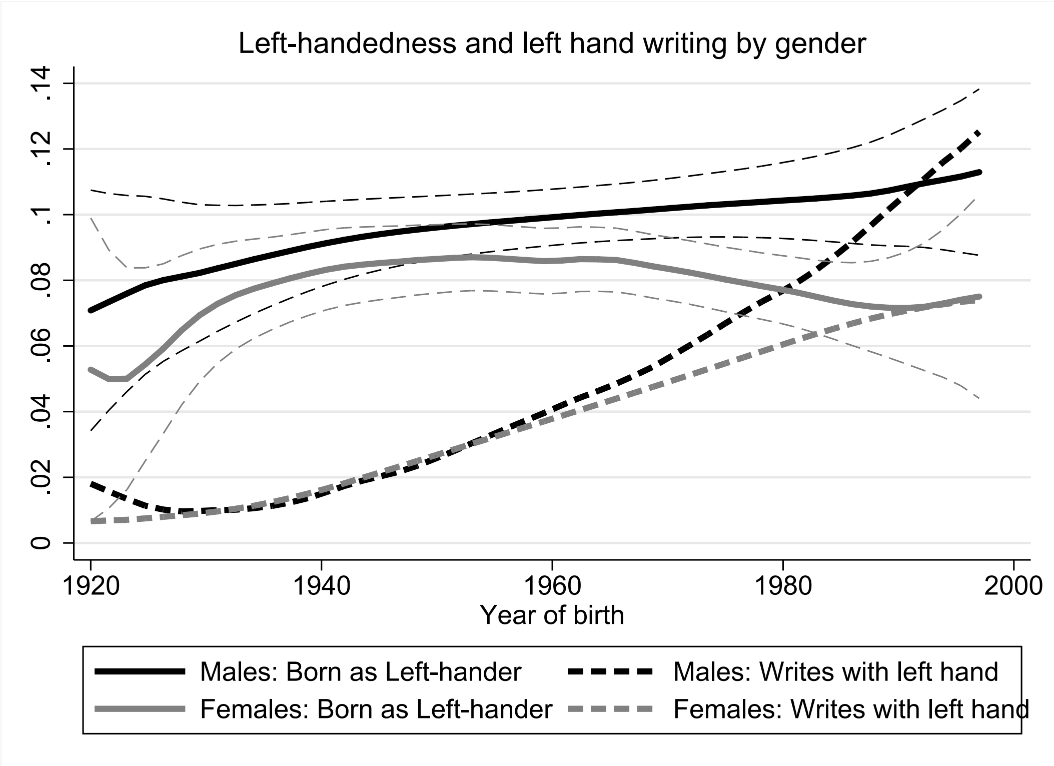
Cohort fixed effects are included in all regressions. Column one of Table 1B.1 demonstrates that females, East Germans, and migrants are significantly less likely to report left-handedness. Interestingly, 2nd generation migrants are not significantly less likely to report left-handedness, although about 50% report having at least one parent born in an Eastern European or Middle Eastern country.

The fact that females are less likely to report left-handedness than males is well-known in the laterality literature (Harris, 1990). The explanations for this phenomenon range from a higher natural predisposition for males, females' increased ability to switch handedness, and stronger social pressure on females to align with norms (Porac et al., 1986; Papadatou-Pastou et al., 2008). Investigating this finding further, Figure 1B.1 repeats Figure 1.2, but splits the sample up by gender. The upper graph in Figure 1B.1 exhibits the level difference between males and females, and it appears that the share of left-handed females actually decreases after the 1960 cohort. However, there exists no significant upward trend among

[19] Germany, Middle East (incl. central Asia), Eastern Europe (incl. Russia), Northern Europe, Southern Europe (Portugal, Greece, Italy, Spain), and other (incl. Africa and Asia). The Middle East includes Turkey, Iran, Syria, Afghanistan, Tunisia, Iraq, Morocco, Kazakhstan, Lebanon, Kirghiztan, Egypt, Tajikistan, Uzbekistan, Azerbaijan, Yemen, Palestine, and Turkmenistan. Eastern Europe includes former Yugoslavia, Romania, Poland, Hungary, Bulgaria, Czech, Russia, Albania, Ukraine, Estonia, Lithuania, Latvia, Croatia, Bosnia, Macedonia, Slovenia, Slovakia, Belarus, Kosovo, Georgia, Serbia, and other Eastern Europe countries. Northern Europe includes Austria, France, Denmark, the UK, Sweden, Norway, Finland, Switzerland, Ireland, Luxemburg, Belgium, and the Netherlands. Other includes all other countries, mainly the USA and Asian and African countries.

males. Linear or quadratic cohort trends in a logit regression for left-handedness among individuals born after 1930 are not significant at any conventional level. The interaction with gender however is negative and significant (p-value 0.054).

Figure 1B.1: Left-handedness and writing by gender and year of birth



Note: Includes 95% confidence intervals for left-handedness.

Although migration decisions are highly endogenous, and might be directly related to left-handedness, I also investigate the heterogeneity of left-hander rates by country of birth (this only applies to 1st generation migrants) in column two. The lower rates among individuals from Middle Eastern and Southern European countries can be explained by religious norms. In Islam, the left hand is the unclean hand, not to be used in human interaction or eating. The Southern European countries are predominantly Catholic (Spain, Italy, Portugal) or Orthodox and considered to be more religious than Northern European countries (Hank and Schaan, 2008). At the same time, the left hand or left side has negative associations in Christianity. Obviously, these cross-country differences are subject to a highly endogenous migration decision, in which left-handedness may play a role. However, the observed cross-country pattern here has been reported in previous studies (Perelle and Ehrman, 1994; Medland et al., 2004). I find no significant differences in left-hander rates between Germans and migrants from its Northern European neighbors, the majority of which come from Austria, France, the United Kingdom, or the Netherlands. Whether there exist differences in religious affiliation is investigated in column three, which excludes migrants.

No significant difference in reported left-handedness exist between Catholics, Protestants, or non-denominational, among German natives (column three). In columns one to three of Table 1B.1, parental education is never a significant predictor of left-handedness and neither is urbanization at age 15. Excluding East Germans in column four does not change any of the previous findings.

If stigmatization against left-handers has been changing over time, it could be the case that certain characteristics predict reported left-handedness in different cohorts. For example, more progressive parents may be more tolerant towards a left-handed child, even when there is discrimination in society as a whole. To investigate this, I split up the sample into four cohort groups of roughly equal size. The first cohort group covers individuals born between 1920 and 1949. These cohorts are most likely to be subject to survival bias, as they lived through World War Two (see Kesternich et al., 2014). Underreporting of left-handedness, as suggested in Figure 1.2, is also most likely to occur in this group. The next three groups comprise respondents born between 1950 and 1960, 1961 and 1970, and 1971 and 1997. However, the overall share of left-handers in all four groups does not differ significantly. Among the left-handed, the share of switched individuals decreases from roughly 91% in the first group, to 83% in the second, 51% in the third, and only 16% in the fourth. Although the enforcement of the right hand writing norm diminishes, the association between parental education and reported left-handedness remains very low, even in the first cohort group. Dummies for parental education are never jointly significant in any sample. The same holds for degree of urbanization. Inclusion of either only maternal or only parental education, and exclusion of urbanization dummies, does not change these findings. This is important, as it suggests there were no significant compositional changes of left-handers with respect to these covariates over time.

The level difference between East and West in the full sample is driven by cohorts born after 1960. The Berlin Wall was built in 1961 and brought a new wave of oppression. As dissidents could no longer simply leave the country, the regime aimed to stigmatize non-conformists to the socialist ideology, starting early in school. Thus, the liberal movement of the 1960s was much less developed in East Germany than in West Germany (Ohse, 2010).

To summarize, I find that left-handedness is only poorly predicted by my covariates, as indicated by the adjusted R^2s of less than 1%. A low correlation of handedness with family background characteristics has been observed by Johnston et al. (2009) in a sample of Australian children and with a broader range of variables. They find no differences between left- and right-handers with respect to either maternal or paternal income, labor force participation, or education. Left-handedness is so nearly random that some studies have employed it as an instrumental variable for the cognitive skills of children (Frijters et al., 2009, 2013).

Furthermore, Goodman (2014) found that perinatal health and maternal handedness are important predictors of left-handedness, while maternal education is not. While I can confirm the latter, my data contain no measure for infant health, such as birth weight. However, I have information on parental handedness for some individuals, because their parents reside in the same household and participated in the grip strength measurement. This sample comprises 1,646 relatively young individuals, who were on average 26 years old at the time of the survey. I find that having a left-handed mother nearly doubles the chance of being left-handed (15.32% vs. 8.48%, p-value 0.007), while the father's handedness is statistically unrelated to own-handedness (9.08% vs. 8.75%, p-value 0.888). These results are robust to controlling for the parent's year of birth and education. Either a left-handed gene is inherited only via the mother, or, more plausibly, children are more keen to use the left hand if they observe their mother doing so, as suggested by Goodman (2014).

Table 1B.1: Regressions on left-handedness

	Linear regression for left-handedness							
	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)
Sample restrictions:			Non-migrants only	Exclude East Germans	By cohort			
					1920-1949	1950-1960	1961-1970	1971-1997
Share of left-handers:	8.95%	8.95%	9.49%	9.21%	8.47%	9.12%	9.45%	8.72%
Female	-0.018***	-0.018***	-0.020***	-0.019***	-0.002	-0.016	-0.016	-0.035***
	(0.005)	(0.005)	(0.007)	(0.006)	(0.011)	(0.011)	(0.011)	(0.010)
East German	-0.016**	-0.015**	-0.010		0.004	-0.006	-0.025*	-0.042***
	(0.007)	(0.007)	(0.009)		(0.013)	(0.014)	(0.014)	(0.014)
Mothers education								
Basic/none	(ref.)							
Middle/other	-0.004	-0.004	-0.005	0.003	0.037	-0.040**	0.014	-0.004
	(0.008)	(0.008)	(0.011)	(0.009)	(0.023)	(0.019)	(0.016)	(0.013)
High	0.007	0.008	0.001	0.012	-0.001	-0.028	0.001	0.022
	(0.012)	(0.012)	(0.015)	(0.013)	(0.034)	(0.026)	(0.024)	(0.018)
Fathers education								
Basic/none	(ref.)							
Middle/other	-0.006	-0.006	-0.003	-0.001	-0.009	0.026	-0.021	-0.005
	(0.008)	(0.008)	(0.011)	(0.010)	(0.022)	(0.022)	(0.016)	(0.013)
High	-0.005	-0.007	-0.003	-0.002	-0.009	0.014	-0.015	-0.011
	(0.009)	(0.009)	(0.012)	(0.011)	(0.023)	(0.021)	(0.019)	(0.016)
Urbanization at age 15								
Large city	(ref.)							
Mid-size city	0.012	0.012	0.012	0.015	0.018	-0.006	0.019	0.012
	(0.009)	(0.009)	(0.012)	(0.010)	(0.020)	(0.019)	(0.019)	(0.017)
Small town	-0.006	-0.006	-0.012	-0.002	0.004	-0.019	-0.014	0.002
	(0.008)	(0.008)	(0.010)	(0.009)	(0.017)	(0.018)	(0.017)	(0.016)
Countryside	-0.011	-0.011	-0.007	-0.008	0.001	-0.007	-0.022	-0.018
	(0.008)	(0.008)	(0.010)	(0.008)	(0.014)	(0.016)	(0.015)	(0.015)
Migration background								
none	(ref.)							
direct	-0.036***			-0.036***	-0.024	-0.030*	-0.038**	-0.052***
	(0.008)			(0.008)	(0.018)	(0.017)	(0.015)	(0.015)
indirect	-0.008			-0.015	0.019	-0.021	-0.013	-0.019
	(0.010)			(0.011)	(0.025)	(0.022)	(0.022)	(0.015)
Country of origin								
Germany		(ref.)						
Middle East		-0.043***						
		(0.013)						
Eastern Europe		-0.036***						
		(0.011)						
Northern Europe		0.020						
		(0.034)						
Southern Europe		-0.067***						
		(0.017)						
Other		-0.024						
		(0.029)						
Religious affiliation								
Catholic			(ref.)					
Protestant			0.006					
			(0.009)					
Denomination free			-0.003					
			(0.010)					
Other			0.019					
			(0.035)					
Cohort fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
N	11,153	11,133	7,341	8,748	2,588	2,624	2,807	3,134
Adjusted R^2	0.002	0.002	0.002	0.002	0.001	0.003	0.000	0.006

Left-handed is a dummy equal to one if the individual self-reports at least once in any survey wave that he is born as a left-hander. Full sample restricted to cohorts between 1920 and 1997. Middle eastern countries include Central Asian countries. Eastern Europe includes Russia. All regressions include cohort fixed effects. Robust standard errors in in parenthesis below. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Chapter 2

Increasing the Credibility of the Twin Instrument*

Abstract: *Twin births are an important instrument for the endogenous fertility decision. However, twin births are not exogenous either as dizygotic twinning is correlated with maternal characteristics. Following the medical literature, we assume that monozygotic twins are exogenous, and construct a new instrument, which corrects for the selection although monozygotic twinning is usually unobserved in survey and administrative datasets. Using administrative data from Sweden, we show that the usual twin instrument is related to observed and unobserved determinants of economic outcomes, while our new instrument is not. In our applications we find that the classical twin instrument underestimates the negative effect of fertility on labor income. This finding is in line with the observation that high earners are more likely to delay childbearing and hence have a higher risk to get dizygotic twins.*

*This chapter is based on joint work with Helmut Farbmacher and Johan Vikstroem.

2.1 Introduction

As fertility decisions are endogenous, most papers on how family size affects maternal and child outcomes use instrumental variable (IV) techniques. One commonly employed instrument are twin births. Early studies that use the twin instrument to study maternal outcomes include Rosenzweig and Wolpin (1980a); Bronars and Grogger (1994); Angrist and Evans (1998); Jacobsen et al. (1999). The twin instrument has also been used to study the prediction of the Becker and Lewis (1973) quantity–quality model, that family size has a negative effect on children's economic outcomes (Rosenzweig and Wolpin, 1980b; Black et al., 2005; Cáceres-Delpiano, 2006; Angrist et al., 2010). Recent applications using the twin instrument include, for instance, Mogstad and Wiswall (2016); Braakmann and Wildman (2016); Lundborg et al. (2017).

However, it has been questioned if having twins—particularly dizygotic twins—really is a random event. In particular, it has been shown that dizygotic twinning depends on, for example, maternal age, height, weight, race, and the use of fertility treatments (Reddy et al., 2005; Fauser et al., 2005).[1] On the other hand, monozygotic (identical) twin births are considered a random event (Tong and Short, 1998; MacGillivray et al., 1988), since they are the result of the random and spontaneous division of a single fertilized egg (e.g., Hall, 2003).[2]

Some studies (Black et al., 2007; Figlio et al., 2014) have already employed the superiority of monozygotic twinning in robustness checks by comparing estimates using all twins as instrument with estimates using only same-sex twins. If the estimates of both instruments are similar in size, this indicates that selection on unobservables is not a problem. However, if the estimates differ, this would cast doubt on the identification strategy. As a response, we construct a new instrument

[1]Some of these variables, such as maternal age and race, are typically observed, while, for instance, fertility treatments, weight and height typically are unobserved in census data.

[2]In a review of the medical literature Bortolus et al. (1999) conclude that it is very rare to find significant correlations between socio-economic characteristics of the parents and monozygotic twin births.

based on monozygotic twins which corrects for the selection bias even though monozygotic twinning is usually unobserved.

Initially, we use longitudinal data from Sweden to show that twin births are correlated with observed and unobserved maternal characteristics and that this correlation is stronger in more recent cohorts. To analyze the selection on unobservables, we use information about pre-pregnancy labor force participation, labor income, and hospitalizations, and conclude that these pre-pregnancy outcomes predict future twin births. This selection is likely to be even more pronounced in data from the US, where twin rates are almost twice as high as those in Sweden. We emphasize, however, that these concerns only apply to dizygotic twin births and not to monozygotic twin births.

We propose a new instrument based on monozygotic twin births which corrects for the non-randomness of twin births. The starting point is the fact that monozygotic twin births are considered to be random events (Tong and Short, 1998; MacGillivray et al., 1988). Our key assumption therefore is that monozygotic twinning is exogenous, but since zygosity rarely is known our approach does not rely on observing zygosity. We show that it is possible to use the observed opposite-sex dizygotic twin mothers to correct the same-sex twin instrument by the remaining selection bias (induced from the same-sex dizygotic twins). This is possible because of the peculiar structure of the data, for instance, since we know that all monozygotic twins are of the same-sex and that dizygotic twin births with same-sex twins are equally likely as dizygotic twins with opposite-sex. Our new approach can easily be implemented using standard regression techniques.

We also discuss ways to relax our main assumption using instead that monozygotic twinning is less endogenous than dizygotic twinning. Here, we add to the growing literature on imperfect instruments by considering misclassified discrete instrumental variables. Ashley (2009) provides the asymptotic distribution of the IV estimator and discusses strategies to assess the robustness of IV inference with imperfect instruments. Nevo and Rosen (2012) examine identification under different assumptions, for instance, that the correlation between the instrument and the error term is less than the correlation between the endogenous variable and the error term. Conley et al. (2012) consider identification and inference for different strategies that use prior information about how close the exclusion restriction is to being satisfied, including also a Bayesian approach. Kraay (2012) and Chan and Tobias (2015) also use a Bayesian approach to capture prior uncertainty about the exclusion restriction.

Our contribution is important for several reasons. Firstly, twin births provide an unexpected fertility shock and twinning usually results in a strong first-stage regression. Secondly, as already mentioned, the twin instrument has been used in several settings, including studies on fertility and maternal outcomes and studies of the child quality-quantity hypothesis.

Thirdly, since the mid-1970's we have seen a rise in the twinning rate, caused by delayed childbearing and an increasing need for fertility treatments (Martin et al., 2012; Fauser et al., 2005). Since the decision to undergo fertility treatment is an endogenous choice, which is clearly affected by the wish or need to postpone motherhood, it is even more likely that the twin births induced by in-vitro fertilization (IVF) are correlated with important socioeconomic characteristics. For instance, Braakmann and Wildman (2016) show that instrumental variables estimates with and without information on fertility treatments might differ substantially in applications to female labor supply and the child quantity–quality relation.[3] This suggests that mothers with twins have become an increasingly selective sample, which poses a threat to the identification of causal effects using the classical twin instrument.

Fourthly, there are only a few other potential variables which can serve as an instrument for endogenous fertility decisions. A commonly used instrument

[3]Moreover, several studies that analyze the quantity–quality trade off explicitly argue that the twin approach is valid because they study cohorts born before the introduction of modern fertility treatments (e.g., Black et al., 2005; Angrist et al., 2010; Åslund and Grönqvist, 2010; Cáceres-Delpiano and Simonsen, 2012).

is parental preference for a mixed sex composition of children. Other previously used instruments for fertility are natural infertility (Agüero and Marks, 2008), successful IVF treatment (Lundborg et al., 2017), and, in cultures with strong son preferences, the sex of the first child (Lee, 2008).

Besides the non-randomness of twin births another concern with the twin instrument, raised by Rozenzweig and Zhang (2009), is that twins have inferior endowments at birth, such as lower APGAR scores and lower birth weight, than singletons. If these differences induce parents to reallocate resources across their children this will violate the exclusion restriction in studies that uses the twin instrument to study quantity-quality effects on non-twin siblings. Rozenzweig and Zhang (2009) find that such differential birth endowment effects are important, while Angrist et al. (2010) find no evidence that would invalidate the exclusion restriction. Another concern with the twin instrument is the close spacing of twins makes their child-rearing more equal, leading to economics of scale in the child quality production. On the other hand, mothers with twins will only have one child-related leave and not two.[4] Note that the assumption that monozygotic twins are at least less endogenous than dizygotic twins is still valid if the birth endowment effect and the economics of scale effect are the same for monozygotic and dizygotic twins. Another important feature of the twin birth instrument is that the composition of compliers can change with time since birth. Mothers who did not get twins can catch up to twin-birth mothers in terms of fertility. This is thoroughly discussed in Braakmann and Wildman (2016). We acknowledge that our new instrument is not able to address this problem.

We use both Swedish and US data to illustrate our new approach. We revisit the study by Angrist and Evans (1998) and use their data on mothers from the 1980 US census. One result is that both the classical twin instrument and the same-sex twin instrument underestimate the true negative effect of fertility on

[4]In many countries, twin parents have some extra months of leave. In Sweden, twin parents have currently three additional months of leave with income related benefits.

labor earnings. This confirms that dizygotic twin mothers are a positively selected sample, partly because high earners are more likely to delay childbearing and hence have a higher risk to get twins. We obtain similar results using Swedish register data both for mothers who got their first child before the strong rise in fertility treatments and for mothers who got their first child during later periods with substantially higher twin rates.

We proceed as follows: Section 2 introduces the Swedish administrative data set and shows the relation of twin's zygosity with observed and, using a panel approach, unobserved maternal characteristics. Section 3 outlines our identification strategy and how it is applied in practice. The two empirical applications are given in Section 4 and Section 5 concludes.

2.2 Zygosity and selection on (un)observables

2.2.1 Data

We use Swedish register data to assess the importance of selection on observable and unobservable variables. The multi-generational register links individuals to their biological parents and contains information on the year and month of birth, which we use to construct information on twin births. The population register contains yearly information on labor income, labor force participation and education. The National Patient Register provides information on all episodes of in-patient care in Sweden. Our sample comprises all mothers who got their first child in the years 1987 to 2006, which gives us roughly 45,000 women per year. Table 2.1 gives some descriptive statistics of our data set over time. To observe a sufficient number of twins, we split the observational period into five cohorts each containing four years. For instance, maternal age at first birth was around 26.3 in the earliest period and 29.2 in the latest, reflecting the well-documented delay in childbearing.

Table 2.1: Summary statistics for our sample of Swedish mothers

	1987–1990		1991–1994		1995–1998		1999–2002		2003–2006	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
# mothers	175,011		174,121		142,083		148,603		167,258	
Socioeconomic characteristics										
Age (at first birth)	26.278	4.4761	26.972	4.6268	27.864	4.7152	28.608	4.7601	29.221	4.8397
Less than nine years of schooling	0.0043	0.0654	0.0051	0.0712	0.0058	0.0760	0.0045	0.0669	0.0058	0.0759
Nine years of schooling	0.2130	0.4094	0.1552	0.3621	0.1274	0.3334	0.1280	0.3341	0.1014	0.3019
Two year high school	0.4474	0.4972	0.4185	0.4933	0.3378	0.4730	0.1926	0.3943	0.0962	0.2949
Three year high school	0.1336	0.3402	0.1758	0.3806	0.2246	0.4173	0.2963	0.4566	0.3230	0.4676
University or college $<$ 3 years	0.1250	0.3307	0.1522	0.3592	0.1879	0.3906	0.1794	0.3837	0.1479	0.3550
University or college $\geq$ 3 years	0.0753	0.2639	0.0918	0.2887	0.1145	0.3184	0.1961	0.3970	0.3201	0.4665
Phd education	0.0012	0.0346	0.0014	0.0374	0.0020	0.0447	0.0031	0.0556	0.0056	0.0746
Pre-pregnancy outcomes (two years before first birth)										
Labor force participation	0.9726	0.1633	0.9570	0.2027	0.8883	0.3150	0.9052	0.2929	0.9169	0.2760
Log labor income	11.450	0.8724	11.523	0.9399	11.382	1.1886	11.562	1.1740	11.720	1.1213
Hospitalization	0.1183	0.4492	0.1140	0.4496	0.0992	0.4331	0.0811	0.3818	0.0819	0.3904
Twin indicators										
Twins ($\ddot{z}$)	0.0103	0.1010	0.0149	0.1210	0.0181	0.1334	0.0193	0.1376	0.0156	0.1238
Same-sex Twins ($\dot{z}$)	0.0073	0.0853	0.0099	0.0989	0.0117	0.1075	0.0119	0.1085	0.0103	0.1007

Notes: Labor income is in SEK. Hospitalization is an indicator for at least one in-patient care episode.

2.2.2 Twin births in Sweden and the US

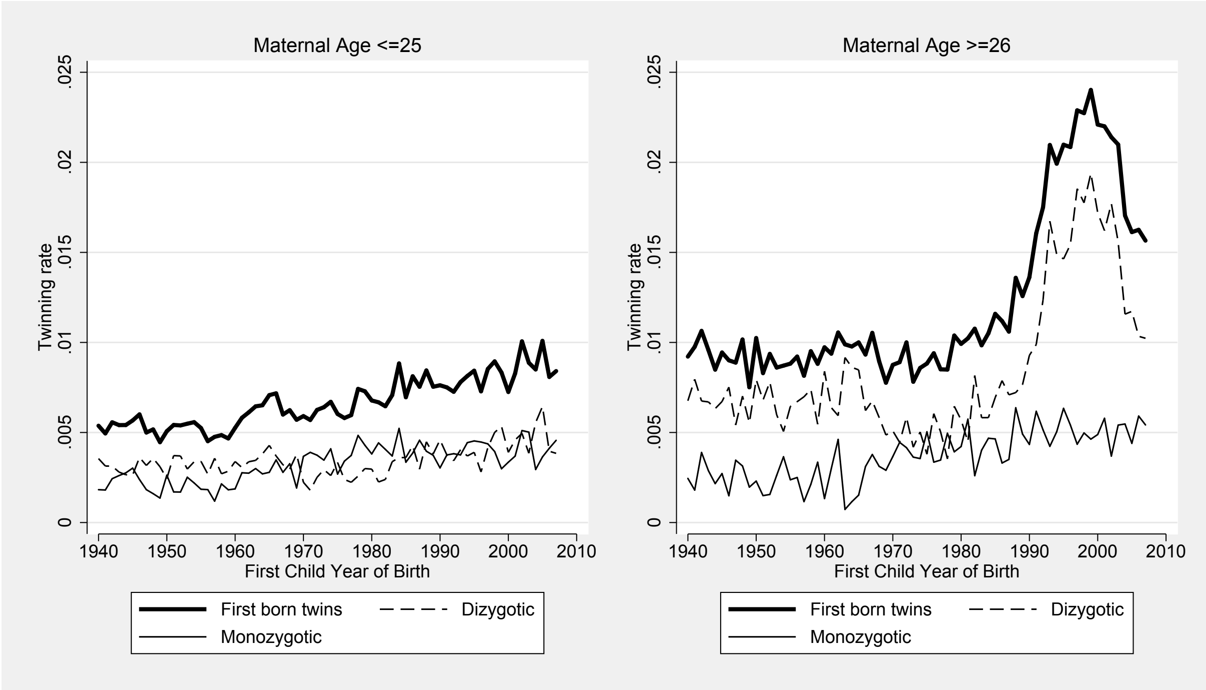
To investigate the changes of twinning in Sweden over time, Figure 2.1 shows the twin rates across the first child's year of birth (separately for younger and older women).[5] The overall twinning rate remains fairly constant between 1950 and 1980 but increases thereafter. While the steady but mild rise in the twin rate of younger mothers from 1980 onwards can be attributed to delayed child bearing, the steep increase in the twin rate of older mothers since 1990 mainly follows the availability of IVF. The drop after 2003 is caused by a recommendation of the Swedish National Board of Health and Welfare regarding the method of elective single embryo transfers (SET), which proceeds by implanting one fertilized egg at a time, instead of several eggs at once, as was done before (Bergh, 2005).

As can be seen in Figure 2.1 (older mothers), the earliest time period (i.e., 1987–1990) we are investigating was just at the beginning of a strong rise in overall twin births (thick solid line). IVF was rather unusual at this time. The later time periods (e.g., 1999–2002), however, are associated with substantially higher twin rates, which are mainly caused by increased fertility treatments. In particular, from 1990 to 2000, the rate of dizygotic twins almost tripled in this age group while the monozygotic rate remained fairly constant in the same period.

The overall twinning rate in the US shows similar patterns to the rate in Sweden, although at a much higher level (see Figure 1A of Kulkarni et al., 2013). The US twin rate (from all parities) was already at 2% in 1980 and increased to more than 3% in 2006. In contrast to Sweden, the US twin rate does not experience the SET related drop and remains high, also by international comparison (Pison and D'Addato, 2006). Thurin et al. (2004) find that twin or higher order pregnancies make up 20% to 25% of all pregnancies induced by IVF in Sweden and Kulkarni et al. (2013) estimate that, in the US, more than one-third of all twins were conceived from fertility assisted pregnancies. Hence, non-random selection

[5] To compute the mono- and dizygotic twinning rates, we apply Weinberg (1901)'s rule, which we discuss in more detail in Section 3.

Figure 2.1: Twin rate in Sweden (firstborn children) between 1940 and 2007.



Note: Statistics based on the Swedish register data described in Section 2.1. To compute the mono- and dizygotic twinning rates, we apply Weinberg (1901)'s rule as described Section 3.

into twinning is likely to be of even more relevance in data from the US.

2.2.3 Selection on observable and unobservable characteristics

Particularly older women need fertility treatments. As postponing childbearing is often related to an individual's labor market decisions, the selection into dizygotic twinning has increased in recent years. Twin mothers are becoming a more and more selected subgroup, which may not be comparable to mothers without twins. For instance, delayed childbearing may help to accumulate more work experience or it may reflect already existing differences in career preferences.

While we can easily determine whether there is any selection on observable characteristics, testing for selection on unobservables is by definition impossible.

However, as many economic determinants are inherently persistent, we can assess the importance of the selection on unobservables by using pre-pregnancy outcomes. That is, we can test whether, conditionally on observable characteristics, twin mothers and non-twin mothers were already different before their first pregnancy. Our pre-pregnancy outcomes are labor force participation, yearly labor income, and hospitalizations two years before the first birth.[6] At this point in time, the future mothers have no children. They might not even know that they will have kids in two years, and they surely do not know that they will have twins. Therefore, the pre-pregnancy outcomes should be causally unaffected by the twin births and the only reason for a pre-pregnancy difference between the twin and non-twin groups is selection on unobserved characteristics.

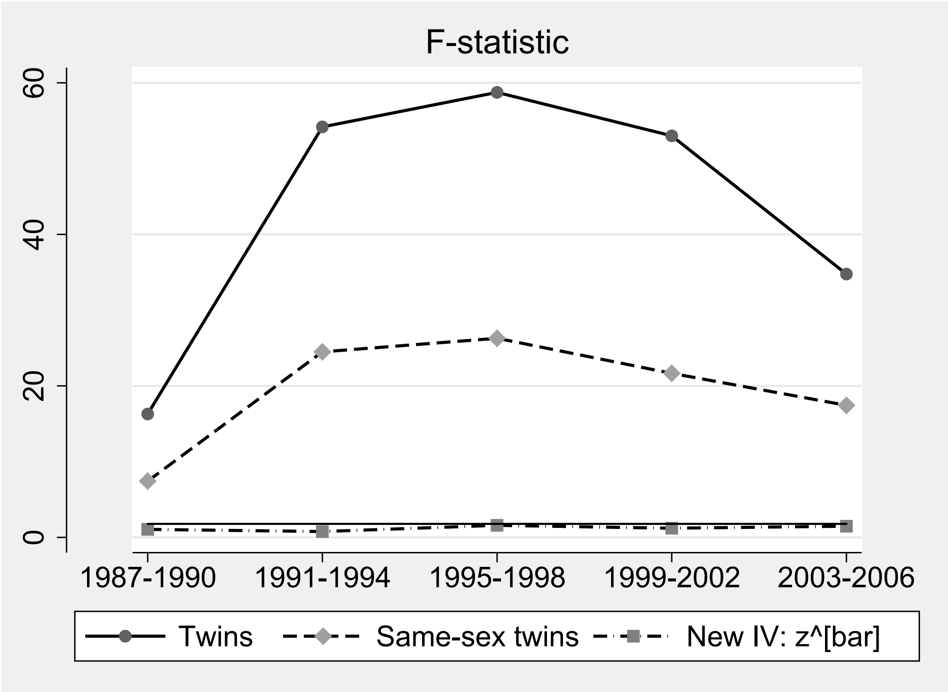
Figures 2.2 and 2.3 show how our observed socio-economic characteristics and the pre-pregnancy outcomes correlate with twin births.[7] Initially, we regress the twin indicators on mother's age at first birth and level of education, and report the overall F-statistic of joint significance (Figure 2.2). The point estimates from these regressions are reported in Table 2C.2, showing, for instance, that the probability of a twin birth is increasing in maternal age.

From the solid line in Figure 2.2 we see that the usual twin indicator correlates with these observables in all time periods. The F-statistic increases strongly from 16.29 in 1987–1990 to over 50 in the later years and does not drop until the year 2003 (for the years 2003-2006 the F-statistic is 34.76). This drop coincides with the rethinking of the SET technique in 2003 to avoid implanting several fertilized eggs at once. The increasing F-statistic reflects the strong rise in twin rates and the increased selection because of fertility treatments and delayed childbearing among mothers with high career preferences. Interestingly, when we use the improved

[6]Labor force participation or employment status is measured in November each year. Labor income includes all cash compensation paid by employers and is based on tax records. For hospitalizations we use an indicator for at least one episode of in-patient care.

[7]Throughout the paper we control for mother's age at first birth using a quadratic polynomial and dummies for mother's education. All results are essentially the same when we use a more flexible regression with age dummies. The results are available from the authors upon request.

Figure 2.2: Assessing the importance of selection on observables.



Note: F-test for joint significance of the regressors from a regression on twin indicators as outcomes and maternal age, age squared and maternal level of education (7 categories) as regressors. The point estimates from the regression are reported in Table C.2. Swedish sample of mothers described in Section 2.1. All models also include year fixed effects.

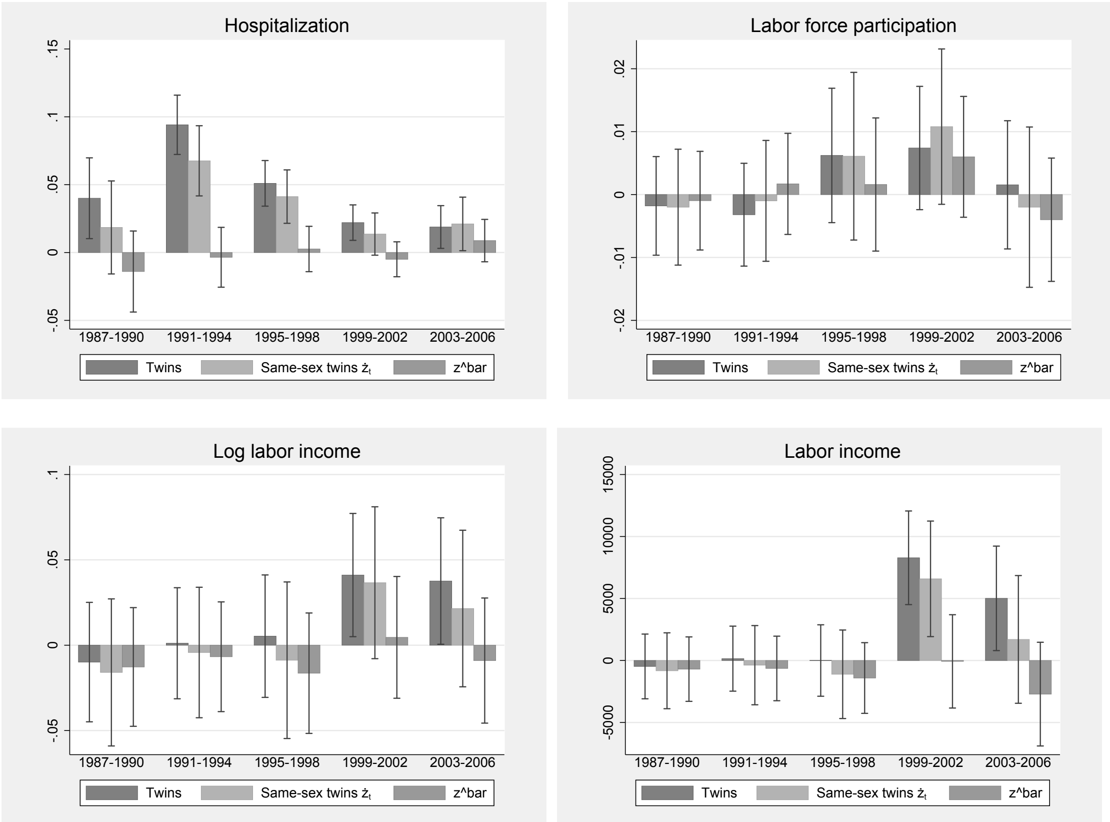
same-sex twin indicator (dashed line), the value of the overall F-test statistic decreases by roughly half in all periods. This indicator variable excludes all opposite-sex twins which cannot be monozygotic.[8,9] Thus, when we exclude twins who have to be dizygotic and thereby implicitly increase the fraction of monozygotic twins, we see a lower dependence of the instrument on socio-economic characteristics.[10]

[8] Opposite-sex twins are not dropped from our analysis. One could in principle also think about dropping the opposite-sex twins but the results should be almost similar due to the low frequency of twinning compared to singleton births.

[9] The F-statistic would decline anyway when the fraction of twins declines, even if we were to randomly exclude some of the twins. To further investigate this relation, we randomly exclude the fraction of opposite-sex twins in a simulation. Using 500 replications, we see, for instance, an average drop of the F-test statistic to 11.71 in the years 1987-1990. This is still distinctly larger than the F-test statistics of 7.43 which we obtain from the regressions on the same-sex twins. This pattern is the same for the other cohorts.

[10] The underlying regressions to Figure 2.2 are reported in Appendix Table 2C.2.

Figure 2.3: Assessing the importance of selection on unobservables.



Note: Differences in pre-pregnancy outcomes (two years before first birth). Note: Estimates and 95%-confidence intervals. Labor income is in SEK. Hospitalization is an indicator for at least one in-patient care episode. All models also include year fixed effects, maternal level of education (7 categories), and a quadratic term in maternal age at birth.

We obtain similar patterns for the pre-pregnancy outcomes. Figure 2.3 shows that—conditionally on the set of covariates—there are still significant differences between women with and without future twins. In the more recent years, women had significantly higher incomes two years before the birth of their twins. The probability of being hospitalized was increased in all cohorts.[11] The significant twin coefficients suggest that there are other (potentially persistent) unobservable variables which may confound estimates based on the conventional definition of

[11] We have information on hospitalization for the period 1987-2005. Therefore, we can include only mothers that gave birth in 1989 or 1990 in the earliest time period sample.

the twin instrument. Similar to the results for selection on observables, these differences become less significant when we use the improved same-sex twin indicator. We will now turn to our methodological contribution and there we will also discuss the remaining results in Figures 2.2 and 2.3.

2.3 Learning from monozygotic twins

In this section, we discuss how the available information about twin births and sibling sex composition can be combined to estimate causal effects even when dizygotic twinning is endogenous and when zygosity is unknown. Consider the model

$$y_i = \beta x_i + u_i, \tag{2.1}$$

where y_i is a scalar denoting the dependent variable, and x_i is the number of children or siblings.[12] The number of children is often used as a unidimensional measure of fertility in labor or health economics (e.g., Angrist and Evans, 1998; Cáceres-Delpiano and Simonsen, 2012), while the number of siblings is used in the literature analysing the child quantity–quality trade-off (Black et al., 2005, 2010). In the former case β is the causal effect of fertility on labor or health outcomes, and in the the latter case β is the causal effect of siblings on, for instance, school performance. The variation in the number of children or siblings is generally considered as endogenous—mainly because having children is a choice and clearly depends on the preferences and socio-economic characteristics of the parents.

Let z^* be an indicator pointing to all mothers with monozygotic twins at their first birth.[13] Monozygotic twinning is most often unobservable—even for the par-

[12] For notational ease, we keep additional explanatory variables implicit. Thus we think about y_i and x_i as variables where the effects of additional explanatory variables have been partialled out, i.e., y_i and x_i are the residuals of a regression of $\tilde{y}_i$ and $\tilde{x}_i$ from a wider model on the additional explanatory variables. In the following, we will suppress the subscript i.

[13] We abstract from higher orders of multiple births such as triplets, since those are very

ents. In one-third of identical twins, each fetus has its own placenta, which is also the case for all dizygotic twins (Bomsel-Helmreich and Al Mufti, 2005). Without further tests, these identical twins cannot be distinguished from fraternal twins unless they have opposite sexes. Typically, neither administrative data sets (like census data) nor surveys contain information on monozygosity. Therefore, we assume having data only on the classical twin indicator, which we denote by $\ddot{z}$, indicating both monozygotic and dizygotic twin births, and an indicator for the sex composition of the first two children (SX). Following Angrist and Evans (1998), the latter variable is defined as $SX = s_1 s_2 + (1 - s_1)(1 - s_2)$, where s_1 and s_2 refer to male first-born and second-born children. Note that SX points to all siblings with the same sex, not only to like-sex twins. Since opposite sex twins can never be monozygotic, we can also define a more precise measure for monozygosity, namely the same-sex twin indicator, $\dot{z}$. Define

$$\begin{aligned}
\dot{z} &= z^* + \dot{e}, \\
\ddot{z} &= z^* + e = z^* + \dot{e} + \ddot{e} = \dot{z} + \ddot{e},
\end{aligned}$$

where e indicates dizygotic twinning, and we allow e to be correlated with the structural error term in Equation 2.1 (i.e., $cov(u, e) \neq 0$). This reflects the clear evidence that dizygotic twinning varies with socio-economic characteristics. Some of these characteristics, such as maternal height and weight, are typically not observed but may have an effect on health or labor outcomes, rendering the classical twin instrument invalid. $\dot{e} = SX \times e$ indicates dizygotic twins with the same sex, and $\ddot{e} = (1 - SX) \times e$ indicates dizygotic twins with a mixed sex composition. Note that $\ddot{e}$ is observable as $\ddot{z} - \dot{z} = \ddot{e}$, while $\dot{e}$ is unobservable for the econometrician, since without further information, same-sex dizygotic twins cannot be distinguished from monozygotic twins.

rare events. In addition these births are the ones that have increased the most due to the IVF availability.

2.3.1 Assumptions

Following the medical literature and the empirical evidence from the previous section, we assume that monozygotic twinning is exogenous or at least less correlated with the structural error term than dizygotic twinning, i.e., the following assumption holds:

***Assumption.* 1.**

monozygotic twinning is "less endogenous" than dizygotic twinning:

$$E(u|z^* = 1) = \theta^* E(u|e = 1) \neq E(u), \text{ with } -1 < \theta^* < 1$$

twinning is relevant:

$$\sigma_{xz^*} \neq 0; \sigma_{xe} \neq 0$$

where z^* is exogenous when the endogeneity parameter $\theta^* = 0$. We also make use of the standard relevance condition of the 2SLS estimator. Since there is an obvious link between having twins and the number of children, relevance is more a technicality ruling out datasets without twin births.

To proceed, we impose two additional assumptions: one medical and one economic. The first assumption is known in epidemiology and medicine as Weinberg (1901)'s differential rule.

***Assumption.* 2.**

Weinberg (1901)'s rule:

$$Pr(\dot{e} = 1) = Pr(\ddot{e} = 1)$$

The rule says that dizygotic twins are equally likely to be of same sex as of opposite sex. The basic assumptions behind this rule are that the probability of a male dizygotic twin (π) is 0.5 (A.2a) and that the sexes in a dizygotic twin set are independent (A.2b). Although the sex ratio at birth is slightly male biased, this rule is generally considered as rather robust (Hardin et al., 2009; Fellman and Eriksson, 2006; Vlietinck et al., 1988; Bulmer, 1976). Nevertheless, in Appendix A we investigate Assumption A.2 using results from the East Flanders Prospective

Twin Survey (EFPTS), and Section 2.3.3 discusses sensitivity analyses with respect to the assumption.

The economic assumption depends on the application one has in mind. It replaces the usual exogeneity assumption of the twin birth indicator, $E(u|\ddot{z} = 1) = E(u|\ddot{z} = 0)$, which is invalid in our setting, by the exogeneity of the sibling sex composition:

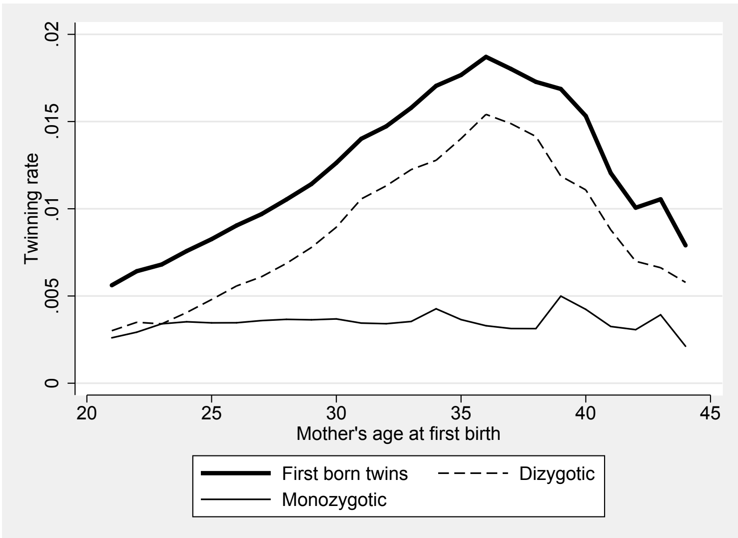
***Assumption.* 3.**
Sex composition of the children is exogenous:

$$E(u|SX = 1, e = 1) = E(u|SX = 0, e = 1) = E(u|e = 1)$$

Assumption A.3 states that the same-sex instrument is exogenous within the group of mothers with dizygotic twins. This is similar to the standard same-sex assumption made by Angrist and Evans (1998), but we argue that our new instrument has several advantages compared to the standard same-sex instrument. First, the same-sex instrument has been criticized because it uses a planned (as opposed to an unplanned) change in fertility to identify the causal effect (Butcher and Case, 1994; Rosenzweig and Wolpin, 2000). Second, the twin instrument could be used to study the effects of having two children instead of only one child, whereas the standard same-sex instrument is only applicable for families with at least two children. For instance, Lundborg et al. (2017) show that the fertility effects differ depending on the margin that is studied. Third, twinning usually results in a strong first-stage regression.

Finally, the external validity of the standard same-sex instrument is debatable, since it identifies the local treatment effect (LATE) for parents that actually have preferences for a mixed sex offspring, and Agüero and Marks (2008) note that these women may differ systematically from the population at large. More recently, Bisbee et al. (2017) use data from 139 country-year censuses to study the external validity of the same-sex instrument, by comparing the actual LATE for one country-year to the extrapolated LATE effect using LATE estimates from other country-years.[14] One conclusion is that the extrapolation works well if it is between similar settings and given that sufficient data is used in the extrapolation. Here, we instead provide additional evidence in favor of the external validity of monozygotic twinning. To this end, Figure 2.4 depicts the twin rates by mothers' age at first birth. According to this figure, monozygotic twinning does not only affect the entire relevant population, but it also seems to be equally likely for women of all age groups.

Figure 2.4: Twin rate in Sweden (firstborn children) by maternal age.



Note: Statistics based on the Swedish register data described in Section 2.1. To compute the mono- and dizygotic twinning rates, we apply Weinberg (1901)'s rule as described Section 3.

One potential threat to Assumption A.3 is that having mixed sex siblings might violate the exclusion restriction. Rosenzweig and Wolpin (2000) argue that same sex siblings could affect the marginal utility of leisure and child rearing costs and, thus, has a direct effect on labor market outcomes. As support of this Rosenzweig and Wolpin (2000) study expenditures per children in rural India, and conclude that the expenditures are lower for same-sex siblings. But, Bütikofer (2010) finds

[14]Specifically, they first characterize the complier populations, and then use these characterizations to extrapolate the LATE estimates from some country-year(s) to another country-year.

no differences in household expenditures for families with different sibling sex composition using data from richer countries (UK and Switzerland). Moreover, the Swedish Household Budget Survey shows that important child rearing costs like clothes and shoes only accounts for about 5.0–6.3 percent of the total household consumption, mainly depending on the number of children (Statistics Sweden, 2010). All this supports Assumption A.3. In Section 2.3.3, we also propose a sensitivity analyses approach to examine the robustness of the estimates with respect to violations of Assumption A.3. This could, for instance, be important if our new approach is applied to data from developing countries.

2.3.2 Identification

In the following, we discuss what can be learned about β using the available information about twinning and the siblings' sex composition. Define β_z^{IV} as the probability limit of the IV estimator for β with z as the instrumental variable. The corresponding estimator is defined as $\hat{\beta}_z^{IV}$. We will use the following notation: σ_{ab} denotes the covariance between any two random variables a and b, and π_a denotes the probability that a binary random variable a is equal to 1. We will also make use of the following relation: If d and w are random variables where d is binary and $E(w) = 0$, then $\sigma_{wd} = \pi_d E(w|d = 1)$.

Using z^* as instrument, we asymptotically get

$$\beta_{z^*}^{IV} \equiv \frac{\sigma_{yz^*}}{\sigma_{xz^*}} \overset{A.1}{=} \beta + \frac{\pi_{z^*}\theta^* E(u|e = 1)}{\sigma_{xz^*}}$$

Although $\hat{\beta}_{z^*}^{IV}$ would be consistent if Assumption A.1 holds and $\theta^* = 0$, it is infeasible since monozygotic twinning is generally unobserved. Estimation based on the observed but misclassified instruments will always be inconsistent, as

$$\beta^{IV}_{\dot{z}} \equiv \frac{\sigma_{y\dot{z}}}{\sigma_{x\dot{z}}} \overset{A.1}{=} \beta + \frac{\pi_{z^*}\theta^* E(u|e=1)}{\sigma_{x\dot{z}}} + \frac{\sigma_{u\dot{e}}}{\sigma_{x\dot{z}}},$$
$$\beta^{IV}_{\ddot{z}} \equiv \frac{\sigma_{y\ddot{z}}}{\sigma_{x\ddot{z}}} \overset{A.1}{=} \beta + \frac{\pi_{z^*}\theta^* E(u|e=1)}{\sigma_{x\ddot{z}}} + \frac{\sigma_{u\dot{e}}}{\sigma_{x\ddot{z}}} + \frac{\sigma_{u\ddot{e}}}{\sigma_{x\ddot{z}}},$$

and because $\sigma_{u\dot{e}} \neq 0$ and $\sigma_{u\ddot{e}} \neq 0$ due to the non-random selection process behind dizygotic twinning. However, using Weinberg's law (Assumption A.2) and the same-sex exogeneity assumption (Assumption A.3), the following result can be derived:

***Proposition* 1.** *If Assumption A.2 and A.3 hold, then*

$$\sigma_{u\dot{e}} = \sigma_{u\ddot{e}}.$$

Proof of P. 1. Note that by the definitions of $\dot{e} = SX \times e$ and $\ddot{e} = (1-SX) \times e$, it follows that $E(u|\dot{e}=1) = E(u|SX=1, e=1)$ and $E(u|\ddot{e}=1) = E(u|SX=0, e=1)$. Furthermore,

$$\begin{aligned}
\sigma_{u\dot{e}} &= \pi_{\dot{e}} E(u|\dot{e}=1) \\
&= \pi_{\dot{e}} E(u|SX=1, e=1) \\
&\overset{A.2}{=} \pi_{\ddot{e}} E(u|SX=1, e=1) \\
&\overset{A.3}{=} \pi_{\ddot{e}} E(u|SX=0, e=1) \\
&= \pi_{\ddot{e}} E(u|\ddot{e}=1) = \sigma_{u\ddot{e}}.
\end{aligned}$$

Q.E.D.

Using Proposition 1 we can derive the following moment condition

$$E(u\bar{z}(\theta)) = 0, \tag{2.2}$$

where $\bar{z}(\theta) = \dot{z} - \lambda(\theta)\ddot{e}$ is a weighted average of two observed variables with $\lambda(\theta) = 1 - \theta(1 - \pi_{\dot{z}}/\pi_{\ddot{e}})$. The moment condition holds if $\theta = \theta^*$, as

$$\begin{aligned} cov(u, \bar{z}(\theta)) &= cov(u, z^* + \dot{e} - \lambda(\theta)\ddot{e}) \\ &= \sigma_{uz^*} + \sigma_{u\dot{e}} - \sigma_{u\ddot{e}} - \theta\tilde{\pi}\sigma_{u\ddot{e}} \\ &\overset{P.1}{=} \pi_{z^*}E(u|z^* = 1) - \theta\tilde{\pi}\sigma_{u\ddot{e}} \\ &\overset{A.1}{=} \pi_{z^*}\theta^* E(u|e = 1) - \theta\tilde{\pi}\sigma_{u\ddot{e}} \\ &\overset{A.3}{=} \pi_{z^*}\theta^*\sigma_{u\ddot{e}}/\pi_{\ddot{e}} - \theta\tilde{\pi}\sigma_{u\ddot{e}} \\ &\overset{A.2}{=} ((\pi_{\dot{z}} - \pi_{\ddot{e}})/\pi_{\ddot{e}})\theta^*\sigma_{u\ddot{e}} - \theta\tilde{\pi}\sigma_{u\ddot{e}} \\ &= (\theta^* - \theta)\tilde{\pi}\sigma_{u\ddot{e}}, \end{aligned}$$

where $\tilde{\pi} = \pi_{\dot{z}}/\pi_{\ddot{e}} - 1$. The intuitive idea behind this new instrument is that we use the observed opposite-sex dizygotic twin mothers to correct for the selection bias induced by same-sex dizygotic twins and possibly also by monozygotic twins. The correction factor $\lambda(\theta)$ also depends on the degree to which monozygotic twins are endogenous.

In the special case where monozygotic twinning is assumed to be exogenous (i.e., $\theta^* = 0$), we get the new instrument $\bar{z}(0) = \ddot{z} - 2\ddot{e}$ by simply subtracting the opposite-sex twins ($\ddot{e}$) twice from the classical twin instrument ($\ddot{z}$). By doing so, we remove not only the endogeneity from the opposite-sex twins but also the endogeneity from the same-sex dizygotic twins.

Assuming that monozygotic twinning is at least less correlated with unobserved characteristics than dizygotic twinning (i.e., $-1 < \theta^* < 1$), we can obtain a set of estimates for β under different assumptions about the degree of endogeneity of monozygotic twinning. For this we construct $\bar{z}(\theta)$ for a grid of values of the endogeneity parameter θ (in between -1 and 1) and calculate the 2SLS estimate separately for each of these variables. This procedure is similar to the idea of imperfect instruments in Nevo and Rosen (2012). They argue that if z is less endogenous than x, the ratio of the correlations between z and u and between x

and u must be between zero and one, i.e., $\lambda = \rho_{zu}/\rho_{xu} \in (0, 1)$. Knowledge of λ would enable the construction of an exogenous instrument, but in its absence, one can use any reasonable value or a set of values between zero and one to construct new instruments and to bound the causal effect.

The set of estimates can be tightened if the selection on observables is informative about the selection on unobservables. For instance, we may get tighter bounds by assuming that selection on unobservables is not an issue as long as the selection on observables is not significant. Following this argument, we could even point-identify β by assuming that there is no selection on unobservables at the value of θ which minimizes the selection on observables. This idea is similar to the approach of Altonji et al. (2005), who also use selection on observables to infer on the selection on unobservables. In a similar way, we assume that the θ which minimizes the correlation between the instrument $\bar{z}(\theta)$ and the observed covariates, also minimizes the correlation between the instrument and the unobservable characteristics. A sufficient condition for this would be that we observe a random subset of all determinants of the outcome variable. In practice, one could use the overall F-statistic of joint significance to measure the selection on observables.

We now return to the results on the selection on observables in Figure 2.2 for all cohorts to assess how the new instrument correlates with mothers' observed characteristics. While the F-statistic of the classical twin instrument and the same-sex twin instrument resembles the IVF-induced twin birth boom depicted in Figure 2.1, the F-statistic of our proposed instrument is between 0.78 and 1.59, and is never significant. This indicates that the observables cannot explain the variation in our new instruments. Turning to the remaining results in Figure 2.3, we find that selection on unobservables is reduced as well, in particular in the more recent cohorts. The proposed instrument is never correlated with the pre-pregnancy outcomes.

2.3.3 Sensitivity analyses

We now propose a sensitivity analyses approach with respect to violations of Assumptions A.2 and A.3. It turns out that a violation of both assumptions can be captured in the same framework. Assuming $\theta^* = 0$ (i.e., monozygotic twinning is exogenous), any violations of Assumption A.2 and A.3 only affect Proposition 1. A generalized version of Assumption A.2 is:

A. 2 g.
Generalized Weinberg (1901)'s rule: $Pr(\dot{e} = 1) = Pr(\ddot{e} = 1)\left(\frac{1}{2\pi(1-\pi)} - 1\right)$

Assumption A.2 is a special case of A.2g with $\pi = 0.5$, where π denotes the probability of a male dizygotic twin. It is also possible to generalize Assumption A.3:

A. 3 g.
Generalized version of sex composition is exogenous: $E(u|\dot{e} = 1) = \gamma\, E(u|\ddot{e} = 1)$

In Assumption A.3, we achieve identification by setting $\gamma = 1$. If the outcome of interest is maternal labor supply, $\gamma > 1$ implies that mothers with dizygotic twins with a mixed sex composition ($\ddot{e}$) on average have lower labor supply than the the mothers with same-sex dizygotic twins ($\dot{e}$). One reason for this could be complementarities of raising children of the same sex, possibly leading to higher maternal labor supply. Under Assumption A.2g and A.3g, Proposition 1 changes to

$$\begin{aligned}
\sigma_{u\dot{e}} &= \pi_{\dot{e}} E(u|\dot{e} = 1) \\
&\overset{A.2g}{=} \left(\frac{1}{2\pi(1-\pi)} - 1\right) \pi_{\ddot{e}} E(u|\dot{e} = 1) \\
&\overset{A.3g}{=} \left(\frac{1}{2\pi(1-\pi)} - 1\right) \pi_{\ddot{e}} \gamma E(u|\ddot{e} = 1) \\
&= \gamma \left(\frac{1}{2\pi(1-\pi)} - 1\right) \sigma_{u\ddot{e}}.
\end{aligned}$$

Interestingly, both violations change Proposition 1 in a multiplicative way. We therefore analyze the robustness of our estimates with respect to γ. To do this we need to have information on plausible values of γ. For violations of Weinberg's rule, we have that the probability of a male dizygotic twin roughly is 0.5144 [99%-CI=(0.5009;0.5279)], according to the East Flanders Prospective Twin Survey (described in Appendix A). This 99%-confidence interval of π would correspond to $\gamma \in [1.000; 1.006]$. For violations of the same-sex assumption, it is more difficult to assess the range of plausible values of γ. In the two applications we will apply a conservative approach and show results for a 20% violation of A.3 (i.e., $\gamma \in [0.8; 1.2]$).

2.4 Empirical applications

2.4.1 Swedish register data

We now apply our new instrument to the Swedish cohort-based samples introduced in Section 2. Our outcome variables are labor force participation and yearly labor income one year after the birth of the first child. We are interested in the effects of having more than one child one year after the first birth and use either the classical ($\ddot{z}$), the same-sex ($\dot{z}$) or our ($\bar{z}$) twin indicator as instruments.[15] If, as we have demonstrated in Section 2, twinning is more endogenous in the recent years than in the earliest cohort, we expect the 2SLS coefficients obtained by using our new instrument to differ more markedly from those obtained by using the classical or the same-sex twin instrument in the recent years. We control for mothers' age at first birth and education, as well as time (year) fixed effects.[16] Note that, within

[15]Sample sizes differ from Section 2 because there mothers had to be working two or more years before their first birth to show up in the register data while here they only need to be working one year after their first birth.

[16]Mothers education is taken from the year of their first birth. If this was missing, we use the information from up to seven subsequent years. As the sample sizes in Table 2.2 indicate, there was a strong birth decline during the late 1990s.

the 1987-1990 cohort, about 10% of the mothers had more than one child the year after the first birth, while this figure is 8% for the 2003–2006 cohort.

Table 2.2 reports estimates for labor force participation, Table 2.3 for labor income and Table 2.4 for log-labor income. Apart from the log-labor income regression in the 1995-1998 cohort, our new instrument gives the strongest effect out of all IV regressions, correcting for positively selected mothers with dizygotic twins. In the earlier two cohorts, the correction is not that important. For example, for the earliest cohort, the estimated effect on labor force participation (Table 2.2) is -6.0% when using the classical twin instrument and -9.4% when using the new instrument—a relative difference of more than half. The results are similar for the years 1991-1994. For two of the three more recent cohorts, the correction is even stronger with a relative difference by a factor of around 1.5. For instance, for the middle cohort (i.e., 1995-1998), the estimated effect on labor force participation is -5.8% with the usual twin instrument and -13.6% with our new instrument. These are economically relevant differences. On the other hand, we also observe that the standard errors are two to three times as large with our new instrument. In the Appendix B, we further investigate the statistical relevance of these differences. The corresponding OLS estimate is -7.4% in the earliest cohort and -9.6% in the most recent cohort. The table also reports the first-stage F-statistics.[17]

Tables 2.3 and 2.4 report estimates for labor income and log-labor income. The pattern for log-labor income (Table 2.4) is similar to the effects on labor force participation, but with a lower magnitude.[18] Table 2.3 contains estimates for labor income in levels, which comprises the effect of fertility on the extensive and intensive margin. Again, the largest difference between the old and the new instrument can be seen in the 1995-1998 cohort, where the effect on labor force

[17]Note that the first-stage F-statistic seems extremely large for the $\ddot{z}$ and $\dot{z}$ instruments, which comes from the fact that as we are looking at short run outcomes only one year after first birth, about 10% (19%) of all mothers that have more than one child gave birth to twins in the 1987–1990 (2003–2006) cohort.

[18]The estimates of the log-income regression in the 1995-1998 cohort indicate a negative selection. As all the other results point to a positive selection, we regard this as an outlier.

participation clearly outweighs the effect on log-income.

Table 2.2: Effect of having more than one child one year after birth on Labor Force Participation - Swedish data

	OLS	2SLS		
		$\ddot{z}$	$\dot{z}$	$\bar{z}$
First child born between 1987 and 1990 (N=184,587)				
More than one child	-0.074***	-0.060***	-0.070***	-0.094***
	(0.003)	(0.010)	(0.012)	(0.023)
First stage F-statistic		1,507,227	1,449,942	414
First child born between 1991 and 1994 (N=182,748)				
More than one child	-0.087***	-0.072***	-0.079***	-0.101***
	(0.004)	(0.009)	(0.012)	(0.028)
First stage F-statistic		1,586,409	1,498,477	337
First child born between 1995 and 1998 (N=149,872)				
More than one child	-0.117***	-0.058***	-0.076***	-0.136***
	(0.004)	(0.010)	(0.011)	(0.031)
First stage F-statistic		1,775,343	1,591,219	249
First child born between 1999 and 2002 (N=158,229)				
More than one child	-0.105***	-0.028***	-0.032***	-0.045
	(0.004)	(0.008)	(0.010)	(0.034)
First stage F-statistic		1,936,390	1,717,190	179
First child born between 2003 and 2006 (N=178,718)				
More than one child	-0.096***	-0.041***	-0.057***	-0.105***
	(0.004)	(0.009)	(0.011)	(0.027)
First stage F-statistic		2,071,708	1,891,536	323

Notes: Each cell reports estimates from one separate regression. OLS and 2SLS estimates using the Swedish data described in Section 2.1. Outcome is an indicator for Labor Force Participation. $\ddot{z}$ is an indicator equal to one if the mother gave birth to twins at first birth, $\dot{z}$ indicates same-sex twins at first birth and $\bar{z}$ is our new twin instrument. Control variables are mothers' education (7 dummies), a quadratic polynomial of age at first birth, and year fixed effects. Robust standard errors in parentheses. ***,**,* indicate significance at the 1%, 5%, and 10% level, respectively.

Table 2.3: Effect of having more than one child one year after birth on Yearly Labor Income - Swedish data

	OLS	2SLS		
		$\ddot{z}$	$\dot{z}$	$\bar{z}$
First child born between 1987 and 1990 (N=184,587)				
More than one child	-15,736***	-15,843***	-16,507***	-18,068***
	(279)	(992)	(1,113)	(2,347)
First stage F-statistic		1,507,227	1,449,942	414
First child born between 1991 and 1994 (N=182,748)				
More than one child	-12,567***	-15,002***	-15,786***	-18,115***
	(317)	(937)	(1,088)	(2,803)
First stage F-statistic		1,586,409	1,498,477	337
First child born between 1995 and 1998 (N=149,872)				
More than one child	-16,748***	-18,398***	-19,410***	-22,838***
	(479)	(1,223)	(1,217)	(4,209)
First stage F-statistic		1,775,343	1,591,219	249
First child born between 1999 and 2002 (N=158,229)				
More than one child	-16,952***	-17,679***	-18,707***	-23,016***
	(603)	(1,463)	(1,747)	(6,227)
First stage F-statistic		1,936,390	1,717,190	179
First child born between 2003 and 2006 (N=178,718)				
More than one child	-17,936***	-18,334***	-19,975***	-25,024***
	(623)	(1,558)	(1,816)	(4,805)
First stage F-statistic		2,071,708	1,891,536	323

Notes: Each cell reports estimates from one separate regression. OLS and 2SLS estimates using the Swedish data described in Section 2.1. Outcome is yearly labor income. $\ddot{z}$ is an indicator equal to one if the mother gave birth to twins at first birth, $\dot{z}$ indicates same-sex twins at first birth and $\bar{z}$ is our new twin instrument. Control variables are mothers' education (7 dummies), a quadratic polynomial of age at first birth, and year fixed effects. Robust standard errors in parentheses. ***,**,* indicate significance at the 1%, 5%, and 10% level, respectively.

Table 2.4: Effect of having more than one child one year after birth on Log(Yearly Labor Income) - Swedish data

	OLS	2SLS		
		$\ddot{z}$	$\dot{z}$	$\bar{z}$
First child born between 1987 and 1990 (N=158,827)				
More than one child	-0.452***	-0.462***	-0.467***	-0.479***
	(0.011)	(0.037)	(0.044)	(0.090)
First stage F-statistic		1,448,829	1,393,111	306
First child born between 1991 and 1994 (N=137,136)				
More than one child	-0.386***	-0.502***	-0.518***	-0.570***
	(0.013)	(0.035)	(0.043)	(0.112)
First stage F-statistic		1,344,403	1,239,239	212
First child born between 1995 and 1998 (N=111,983)				
More than one child	-0.391***	-0.501***	-0.463***	-0.315**
	(0.017)	(0.034)	(0.042)	(0.134)
First stage F-statistic		1,625,512	1,394,021	136
First child born between 1999 and 2002 (N=124,239)				
More than one child	-0.367***	-0.475***	-0.484***	-0.521***
	(0.016)	(0.032)	(0.041)	(0.141)
First stage F-statistic		1,812,816	1,551,194	129
First child born between 2003 and 2006 (N=139,171)				
More than one child	-0.230***	-0.294***	-0.280***	-0.311***
	(0.014)	(0.032)	(0.040)	(0.108)
First stage F-statistic		1,939,387	1,749,103	203

Notes: Each cell reports estimates from one separate regression. OLS and 2SLS estimates using data from the Swedish data described in Section 2.1 Outcome is log yearly labor income using the sample of mothers with non-zero labor income. $\ddot{z}$ is an indicator equal to one if the mother gave birth to twins at first birth. $\ddot{z}$ is an indicator equal to one if the mother gave birth to twins at first birth, $\dot{z}$ indicates same-sex twins at first birth and $\bar{z}$ is our new twin instrument. Control variables are mothers' education (7 dummies), a quadratic polynomial of age at first birth, and year fixed effects. Robust standard errors in parentheses. ***,**,* indicate significance at the 1%, 5%, and 10% level, respectively.

2.4.2 1980 US Census data

To illustrate the broad usefulness of our approach, we investigate its relevance using a second application. We revisit the study by Angrist and Evans (1998), in short AE hereafter. The sample consists of all (married and unmarried) mothers aged 21 to 35 with at least two children from the 1980 US census.[19] We use age, age at first birth, sex of the first/second child, and dummies for being black, Hispanic, or of another race as covariates. For a detailed description of the variables, we refer to AE, Table 2.

AE use the usual twin indicator ($\ddot{z}$) and an indicator for same-sex siblings. To this we add the same-sex twin indicator ($\dot{z}$) and our two new instruments. $\bar{z}(0)$ is constructed by assuming that $\theta = 0$, i.e., monozygotic twins are uncorrelated with the structural error term. In practice, this delivers an instrument which takes on a value of -1 for all opposite-sex twins, a value of 1 for same-sex twins, and 0 for non-twin mothers. To construct $\bar{z}(\theta_{min})$, we derive θ_{min} as the θ which minimizes the overall F-statistic in a regression of $\bar{z}(\theta)$ on the covariates. A grid search delivers $\theta_{min} = -0.07$ for the whole sample and $\theta_{min} = 0.01$ for the sample of working mothers.

We study the effects of having more than two children on annual labor income using our various instruments. The covariates are the same as in AE, but our sample size is 394,840 instead of 394,835. The first panel of Table 2.5 shows results for labor income as the dependent variable. Column 1 reports a highly significant negative effect of -3,762 on annual labor income when we use OLS to estimate the fertility parameter. Using twins as instrument this coefficient reduces to -1,228 (column 2). We also find large differences between the estimated effects using the usual instruments and those from using our new twin instruments. The absolute size of the coefficients increases with the share of monozygotic twin mothers in the instrument. The effect is lowest when using all twins (-1,228), but

[19] AE restricted their analysis using twins to data from the 1980 US census, which allows us to reliably identify twins using quarter of birth.

almost doubles (-2,465 and -2,333) when using $\bar{z}(0)$ or $\bar{z}(\theta_{min})$. The increase in the coefficients indicates that dizygotic twin mothers are a positively selected sample, which lead to an underestimation of the true effect. This was to be expected from the known relation between maternal characteristics (particularly, maternal age) and dizygotic twinning. For instance, women who earn more and/or have higher career preferences may also be more likely to postpone motherhood, which would increase the likelihood of dizygotic twinning. The estimate for labor income using the same-sex instrument of -1,902 is in between the estimates from the twin instruments. The different effect size can be attributed to the identification of different local average treatment effects (Angrist et al., 1996). Note that the first stage F-statistic of 632 and 855 of the new instruments are much lower than those of the usual twin instruments $\ddot{z}$ and $\dot{z}$, but are still clearly above the rule of thumb value of 10 (Staiger and Stock, 1997).

The last row in the first panel of Table 2.5 reports the F-statistics of regressions of each respective instrument on the covariates to assess the importance of selection on observables in the US data. The overall F-statistic decreases from 9.45 to 3.85 when using the improved same-sex twin instrument, as compared to the overall twin instrument. As in the Swedish data, our new instruments are the least correlated with the mothers' observable characteristics. Although there still seem to be small correlations with mother's age and age at first birth, the overall F-statistics of 1.13 and 1.11 are insignificant.

Panel two and three of Table 2.5 report results for the probability of working and log-labor income. In the latter case we exclude mothers with zero earnings. For the probability of working we see small differences between the different IV estimates. For log-labor income we see a similar pattern as for labor income, with larger labor supply effects when using our new instruments ($\bar{z}$) compared to the usual twin instrument ($\ddot{z}$) and the same-sex twin instrument ($\dot{z}$). Remember that the primary reason why we would expect a different estimate from the conventional twin instrument and our new instrument is that dizygotic mothers are positively selected, with respect to -among many other variables- career preferences. Thus, our results for 1980 indicate that the unobserved heterogeneity relating to the extensive labor supply margin (probability of working) is limited, while the unobserved heterogeneity relating to the intensive margin (log-labor income) is more substantial.

Finally, we make a detailed comparison of the results from the two applications. Note that AE use cross-sectional census data from 1980 where fertility and outcome variables are only observed in that year, while in the Swedish application outcomes are observed one year after first birth. For comparison reasons we also construct a Swedish sample in similar way as the AE census data, using fertility and labor market outcomes in 1990 and applying the same sample restrictions as in AE. In line with AE, the endogenous variable of interest is now an indicator equal to one if the mother has more than two children. With this sample we use the twins instruments and the same-sex instrument. The results for the 1990 Swedish "census" in Table 2C.1 reveal smaller labor supply and income effects for Sweden compared to the US data.[20] The difference in IV estimates for labor income in levels and logs are, however, not very different across the twin-based instruments. The effect on income using the same-sex instrument (-23,765 SEK, i.e., roughly -$4,021) are considerable larger not only compared to the estimates using our new instrument (-7,544 SEK, i.e., roughly -$1,276) but also in comparison to the corresponding estimates in the US data (-$1,902). This may point to a rather special complier group for the same-sex instrument in Sweden.

[20]Note that the average exchange rate was 5.91 SEK to 1 USD in 1990.

Table 2.5: Effect of having more than two children - US Census data

	OLS	2SLS				
		$\ddot{z}$	$\dot{z}$	$\bar{z}(0)$	$\bar{z}(\theta_{min})$	Same-sex
Yearly Labor Income (N=394,840)						
More than two children	-3,762***	-1,228***	-1,586***	-2,465***	-2,333**	-1,902***
	(34)	(299)	(320)	(738)	(1,000)	(546)
First Stage F		60,239	44,576	632	855	1,675
Selection on obs. (F-stat.)		9.45***	3.85***	1.13	1.11	2.04*
Worked for pay in last year (N=394,840)						
More than two children	-0.176***	-0.081***	-0.082***	-0.084**	-0.084**	-0.117***
	(0.002)	(0.014)	(0.017)	(0.034)	(0.040)	(0.025)
First Stage F		60,239	44,576	632	855	1,675
Selection on obs. (F-stat.)		9.45***	3.85***	1.13	1.11	2.04*
Log(Yearly Labor Income) (N=220,502)						
More than two children	-0.353***	-0.072	-0.112**	-0.215*	-0.217	-0.135
	(0.006)	(0.045)	(0.054)	(0.117)	(0.189)	(0.092)
First Stage F-statistic		35,754	25,484	292	280	841
Selection on observables F-statistic		8.65***	3.53***	0.72	0.72	1.75*

Notes: OLS and 2SLS estimates using data from the 1980 US Census. All models also include age, age at first birth, sex of the 1st child, sex of the 2nd child, and dummies for being black, hispanic, or of other race. Selection on observables F-statistic refers to the F-statistic of a regression of the respective instrument on the above covariates, except sex of the 1st and 2nd child for the same-sex instrument. $\ddot{z}$ is an indicator equal to one if the mother gave birth to twins at first birth, $\dot{z}$ indicates same-sex twins at first birth and $\bar{z}$ is our new twin instrument. θ_{min} equals -0.07 for worked in last year sample and 0.01 in the sample of working mothers. Robust standard errors in parentheses. For the regression with $\bar{z}(\theta_{\min})$ we use a bootstrap with 1000 replications to obtain the standard errors. ***,**,* indicate significance at the 1%, 5%, and 10% level, respectively.

2.4.3 Sensitivity analyses

We now report results from the sensitivity analysis as described in Section 2.3.3. Figures 2C.2 and 2C.1 in the Appendix report results for the Swedish and the US data, respectively. We report results for a 20% violation of A.3g (i.e., $\gamma \in [0.8; 1.2]$). In the US data a 20% violation of A.3 still leave the reported estimates in the labor force participation regression almost unchanged. The point estimates of the wage regression change slightly with γ (getting larger in absolute terms). The estimates remain significant. In contrast to the US results, the Swedish results are very robust in the wage regression and vary slightly in the labor force participation regression. The estimates remain clearly significant.

2.5 Conclusions

Twin births are a popular instrumental variable for the endogenous fertility decision and family size. However, identification of causal effects might fail as having dizygotic twins is strongly related to mothers' age, height, weight, race, and the use of fertility treatments, such as in-vitro fertilizations. To overcome this, we provide a new instrument that corrects for the selection bias introduced by dizygotic twins, even if zygosity is unknown. The new approach depends on a parameter θ, which reflects the researcher's assumption about the strength of the relation between the structural error term and monozygotic twinning, relative to dizygotic twinning.

Although exogeneity is not directly testable, we find supporting evidence for the exogeneity of monozygotic twinning (corresponding to $\theta = 0$). First, we do not find significant correlations between observed covariates and our new instrument. And more importantly, we do not find any significant correlation between our instrument and the lagged dependent outcomes, which contain (time constant) unobservables.

We could, however, also assume that monozygotic twins are not fully exogenous but are less endogenous than dizygotic twins ($\theta \in (-1, 1)$). In this case we propose to set the parameter θ to the value that minimizes the overall F-statistic from a regression of the new instrument on the observed variables under the assumption that the selection on observables is informative about the selection on unobservables. In contrast to the usual instruments (any twins and same-sex twins), we show using Swedish register data that the new instrument is completely unrelated to important pre-pregnancy outcomes.

Additionally, we apply our new approach to both Swedish and US data. Our main finding is that the usual twin instruments underestimate the true negative effect of fertility on labor force participation and earnings. This indicates that twin mothers are a positively selected sample, which is in line with the observation that high earners are more likely to delay childbearing and, hence, have a higher risk to get dizygotic twins.

Appendix

2.A Empirical assessment of Weinberg's rule

East Flanders Prospective Twin Survey (EFPTS) is a population-based registry of multiple births in East Flanders (Belgium). The EFPTS distinguishes itself from other twin registries because the information has been collected by the obstetricians at birth (see Derom *et al.*, 2006 for further information about the EFPTS database). This dataset contains information about the zygosity of the twins, which allows us to test Assumptions A.2a and A.2b.

To investigate the robustness of Assumption A.2a, we can derive a generalized rule which requires only independence (Assumption A.2b) to hold. It is—up to a factor ($f = 1/(2\pi(1-\pi)) - 1$)—equal to Weinberg's differential rule

$$Pr(\dot{e} = 1) = Pr(\ddot{e} = 1)\left(\frac{1}{2\pi(1-\pi)} - 1\right). \tag{2.3}$$

Weinberg's rule is the special case in which $f = 1$. Considering the 99% confidence interval of π from the EFPTS data (99%-CI=[0.5009;0.5279]), the corresponding factor f ranges from 1.000 to 1.006, which makes Weinberg's rule an accurate approximation given that independence (A.2b) holds.

To test whether the sexes in a dizygotic twin set are independent (Assumption A.2b), we also use the EFPTS data. Table 2A.1 shows the observed sex composition of dizygotic twins and the expected frequencies under the null hypothesis of independence. The corresponding χ^2 test statistic is 0.753 (p-value: 0.385), so that independence cannot be rejected.

Table 2A.1: Sex composition of dizygotic twins in the East Flanders Prospective Twin Survey

	girl	boy	
girl	1078 [1063.44]	1112 [1126.56]	2190
boy	1112 [1126.56]	1208 [1193.44]	2320
	2190	2320	4510

Notes: Expected frequencies (under independence) in brackets. Source Derom et al. (2006)

2.B Statistic relevance of the differences in the instruments

To investigate the statistical relevance between two IV estimates using different instruments, we apply two approaches. The first is to bootstrap both estimates, say $\hat{\beta}^{IV}_{\ddot{z}}$ and $\hat{\beta}^{IV}_{\bar{z}(0)}$, B times to get $\hat{\beta}^{IV}_{\ddot{z},b}$ and $\hat{\beta}^{IV}_{\bar{z}(0),b}$ for $b = 1, ..., B$. We then compute the fraction of bootstrap replicates in which the difference between $\beta^{IV}_{\bar{z}(0)}$ and $\beta^{IV}_{\ddot{z}}$ was smaller than zero:

$$\frac{\Sigma_b \mathbb{1}\{\hat{\beta}^{IV}_{\bar{z}(0),b} - \hat{\beta}^{IV}_{\ddot{z},b} < 0\}}{B}.$$

In a second approach we compute the following t-statistic

$$t = \frac{\hat{\beta}^{IV}_{\ddot{z}} - \hat{\beta}^{IV}_{\bar{z}(0)}}{\sqrt{\hat{V}(\hat{\beta}^{IV}_{\ddot{z}}) + \hat{V}(\hat{\beta}^{IV}_{\bar{z}(0)}) - 2\,\hat{\mathrm{cov}}(\hat{\beta}^{IV}_{\ddot{z}}, \hat{\beta}^{IV}_{\bar{z}(0)})}}, \tag{2.4}$$

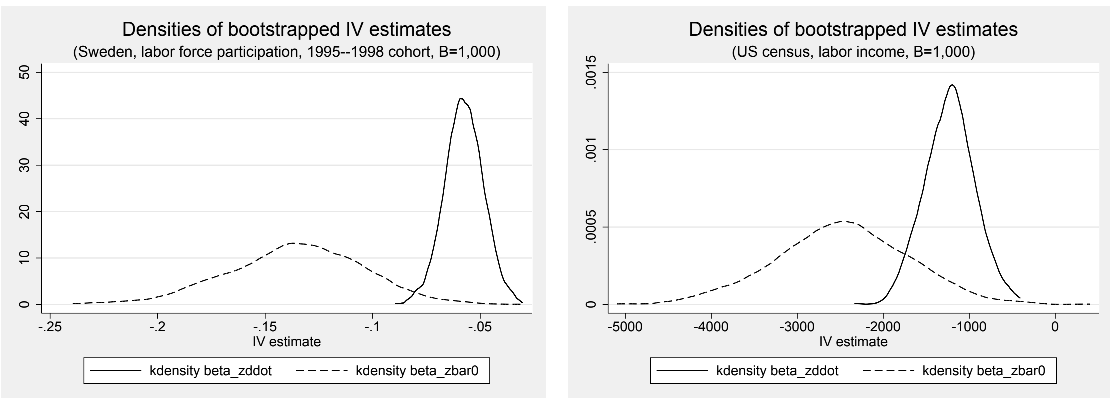
where $\hat{\beta}^{IV}_{\ddot{z}}$ and $\hat{\beta}^{IV}_{\bar{z}(0)}$ are the estimated coefficients on the full sample, $\hat{V}(\hat{\beta}^{IV}_{\ddot{z}})$ and $\hat{V}(\hat{\beta}^{IV}_{\bar{z}(0)})$ are the respective estimated variances from the original IV regressions and $\hat{\text{cov}}(\hat{\beta}^{IV}_{\ddot{z}}, \hat{\beta}^{IV}_{\bar{z}(0)})$ is estimated using 1000 bootstrap replications.

Turning to our applications in Section 4, we found a large difference in IV estimates on labor force participation for the Swedish 1995–1998 cohort when using $\ddot{z}$ or $\bar{z}(0)$ as an instrument (Table 2, panel 3). In the former case, the estimate was -5.8% and in the latter it was -13.6%. While the difference is economically relevant, we now set out to check its statistical relevance. Bootstrapping the estimates 1,000 times, we find that in 998 cases $\hat{\beta}^{IV}_{\bar{z}(0),b}$ was larger in absolute terms than $\hat{\beta}^{IV}_{\ddot{z},b}$. The t-statistic from (2.4) is 2.649 with a p-value of 0.008.

In the AE application (Table 5, panel 1), the IV estimate for having more than two children on labor income was -1,228 when using $\ddot{z}$, while it was -2,465 when using $\bar{z}(0)$. Bootstrapping both estimates 1000 times shows that in the vast majority of cases (940 out of 1000) $\hat{\beta}^{IV}_{\bar{z}(0),b}$ was lower than $\hat{\beta}^{IV}_{\ddot{z},b}$. The t-statistic from (2.4) is 1.631 with a p-value of 0.103.

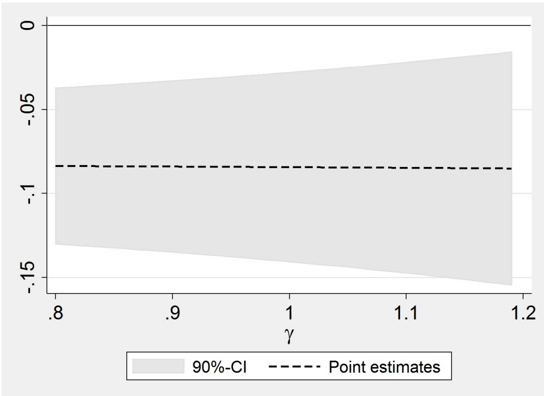
We also noted in Section 4 that estimates from our new instrument have a larger standard error than those from the usual twin IV's. This can also be seen from kernel density graphs over the bootstrapped estimates shown in Figure 2B.1. The estimates are particularly well separated in the Swedish application.

Figure 2B.1: Kernel density estimates of bootstrap replications

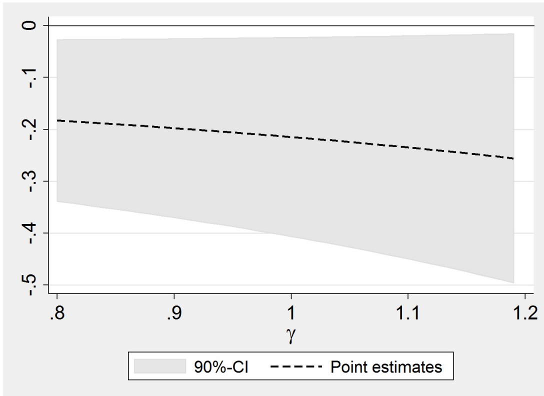


2.C Additional Tables and Figures

Figure 2C.1: Robustness analysis of the effect of having more than two children - US Census data



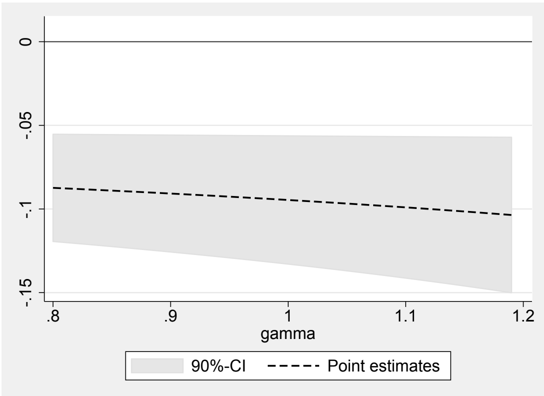
(a) Worked for pay in last year



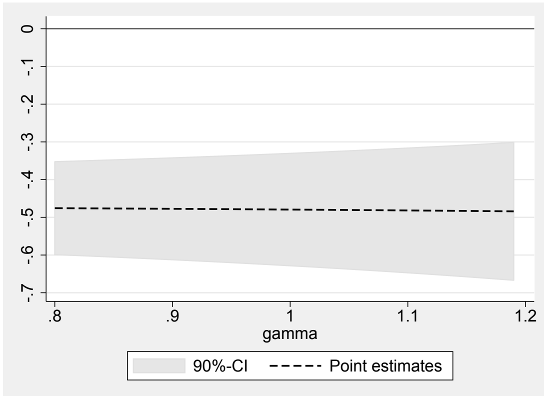
(b) Log(labor income)

Figure 2C.2: Robustness analysis of the effect of having more than two children - Swedish data

Sample: 1987-1990

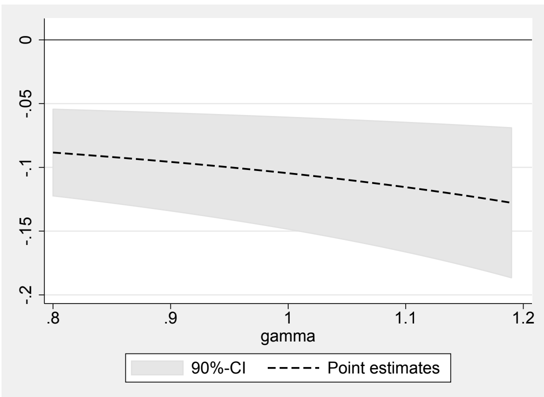


(a) Worked for pay in last year

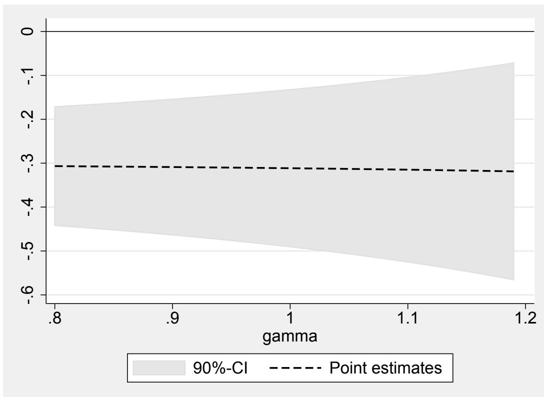


(b) Log(labor income)

Sample: 2003-2006



(c) Worked for pay in last year



(d) Log(labor income)

Table 2C.1: Effect of having more than two children - Swedish data in 1990

	OLS	2SLS				
		$\ddot{z}$	$\dot{z}$	$\bar{z}(0)$	$\bar{z}(\theta_{min})$	Same-sex
Yearly Labor Income (N=287,095)						
More than two children	-28,677***	-7,100***	-7,228***	-7,544**	-7,552*	-23,765***
	(245)	(1,494)	(1,763)	(3,707)	(4,411)	(4,924)
First Stage F		42,738	29,643	511	489	702
Selection on observables F-stat		17.22***	6.95***	1.24	1.24	2.76*
Labor Force Participation (N=287,095)						
More than two children	-0.080***	-0.040***	-0.035***	-0.022	-0.022	-0.039
	(0.002)	(0.009)	(0.011)	(0.023)	(0.030)	(0.028)
First Stage F		42,738	29,643	511	489	702
Selection on observables F-stat		17.22***	6.95***	1.24	1.24	2.76*
Log(Yearly Labor Income) (N=253,551)						
More than two children	-0.465***	-0.082**	-0.079**	-0.072	-0.072	-0.524***
	(0.005)	(0.033)	(0.039)	(0.081)	(0.102)	(0.103)
First Stage F		40,956	28,510	458	438	667
Selection on observables F-stat		14.11***	6.02***	1.27	1.26	2.10*

Notes: OLS and 2SLS estimates using data from Swedish administrative data in 1990. Following AE, the sample is restricted to mothers between age 25 and 35 with at least two children and first child below age 18. All models include dummies for maternal age in 1990, age at first birth and sex of the 1st and 2nd child. The average exchange rate was 5.91 SEK to 1 USD in 1990. Robust standard errors in parentheses. ***,**,* indicate significance at the 1%, 5%, and 10% level, respectively.

Table 2C.2: Assessing the importance of selection on observables. Point estimates. Swedish data

	$\ddot{z}$ (1)	$\dot{z}$ (2)	$\bar{z}$ (3)
Sample: 1987-1990 (N=175,011)			
Maternal age/100	0.113***	0.047***	-0.019
Maternal age^2/100	-0.001***	-0.000***	0.000
Nine years of schooling	-0.007	-0.004	-0.001
High school 2 years	-0.007	-0.004	-0.000
High school 3 years	-0.008	-0.004	-0.000
University or college < 3 years	-0.008	-0.004	-0.000
University or college ≥ 3 years	-0.008	-0.004	-0.000
Phd education	-0.009	-0.005	-0.019
Sample: 1991-1994 (N=174,121)			
Maternal age/100	-0.016	-0.004	0.008
Maternal age^2/100	0.004***	0.002**	0.000
Nine years of schooling	0.008***	0.004***	-0.000
High school 2 years	0.007***	0.003***	-0.000
High school 3 years	0.004***	0.002***	-0.000
University or college < 3 years	0.005***	0.002***	-0.001
University or college ≥ 3 years	0.003***	0.001***	-0.001
Phd education	-0.011***	-0.011***	-0.010
Sample: 1995-1998 (N=142,083)			
Maternal age/100	0.206***	0.059***	-0.087
Maternal age^2/100	-0.000***	0.001***	0.002
Nine years of schooling	0.005**	0.003	0.001
Two year high school	0.004**	0.003	0.003
Three year high school	0.002**	0.002	0.003
University or college < 3 years	0.001**	0.001	0.002
University or college ≥ 3 years	0.002**	0.003	0.005
Phd education	0.010**	0.009	0.007
Sample: 1999-2002 (N=148,603)			
Maternal age/100	0.131***	0.006***	-0.119
Maternal age^2/100	0.001***	0.002***	0.002
Nine years of schooling	-0.007	0.003	0.012
High school 2 years	-0.008	0.002	0.012
High school 3 years	-0.009	0.002	0.013
University or college < 3 years	-0.009	0.002	0.013
University or college ≥ 3 years	-0.010	0.000	0.011
Phd education	-0.015	-0.008	-0.000
Sample: 2003-2006 (N=167,258)			
Maternal age/100	-0.237***	-0.101***	0.034
Maternal age^2/100	0.007***	0.003***	-0.000
Nine years of schooling	-0.002***	-0.001**	0.001
High school 2 years	0.004***	-0.003**	0.002
High school 3 years	-0.000***	-0.000**	-0.000
University or college < 3 years	-0.001***	-0.000**	-0.000
University or college ≥ 3 years	-0.001***	0.000**	0.001
Phd education	0.000***	-0.003**	-0.006

Notes: LPM estimates using the Swedish sample of mothers described in Section 2.1. The outcomes are twin indicators. $\ddot{z}$ is an indicator equal to one if the mother gave birth to twins at first birth. $\dot{z}$ indicates same-sex twins at first birth and $\bar{z}$ is our new twin instrument. All models also include year fixed effects. ***,**,* indicate significance at the 1%, 5%, and 10% level, respectively.

Chapter 3

The Burden of Child Rearing and Working on Maternal Mortality*

Abstract: *In times of increasing female labor market participation and policy efforts to combine work and family life, it is important to understand the consequences of actively raising children and simultaneously pursuing a career for mothers' health. Based on Swedish administrative data we document strongly increased old-age mortality rates among mothers that potentially experienced a double burden. We use twins at first birth as an unplanned shock to fertility and proxy labor force attachment by stratifying the sample by education and pension income. In line with the double burden hypothesis, the effect of having twins is largest among highly educated mothers and those with above-median pension income. Deaths due to lung cancer, COPD and heart attacks, which are strongly associated with stress during life, are over-proportionally increased.*

*This chapter is based on joint work with Tabea Bucher-Koenen, Helmut Farbmacher and Johan Vikstroem.

3.1 Introduction

In times of demographic change and a decreasing work force policy makers in many developed countries aim at increasing female labor supply in order to better tap into hidden reserves. At the same time preventing birth rates from declining further or even increasing them is on the agenda (Jaumotte, 2004). However, actively raising children and pursuing a career are two conflicting, because time consuming, activities. A large literature documents the negative effect of fertility on female labor force participation and hours worked (Angrist and Evans, 1998; Lundborg et al., 2017). Another strand of literature evaluates policy measures that change the incentives for labor force participation and child bearing (Del Boca, 2002; Lalive and Zweimüller, 2009; Schönberg and Ludsteck, 2014). One often made argument in the debate on how to improve the reconciliation of family and working life is that mothers (and fathers) need to be shielded from stress that could occur form the double burden of working and caring for children at the same time. Up to this point, there exists very little evidence on this double burden and the consequences of such a burden for parental health. This paper aims at filling this gap by studying the effects of past fertility shocks and their interaction with labor force activity on health later in life.

In order to study the combined effect of fertility and labor force activity on maternal health, an ideal set-up would provide exogenous variation in both labor market participation and having and raising children. We use twins at first birth as an unplanned shock to fertility. Two potential threats to this approach are that twin births increase with mothers' age at birth and the availability of in-vitro fertilization. We can use this strategy nevertheless, because the cohorts examined in this study had their first children between 1940 and 1970, which is well before any major impact of fertility treatments on the number of twin births. Additionally, we condition on mothers' age at first birth in all our analysis. In order to examine the interaction of fertility and labor market activity, we then stratify the sample along educational attainment and pension income at age 70, respectively. These variables are strongly related to labor force attachment and help us examine the heterogeneity of the effect of having twins on health.

This analysis is based on administrative birth and death registries from Sweden. The sample for our analysis includes more than 400,000 mothers that were 55-65 years old, alive and resident in Sweden in 1990. This means that we analyze mothers with completed fertility histories. We can follow these mothers over time for twenty years until 2010. If they died within our observation window, administrative death certificates give information on the cause of death. We focus on mothers, because women are significantly more likely than men to find themselves in a situation where family and working life are in conflict—at least for the cohorts of women examined here. In addition to their fertility history we can draw information on these womens' sociodemographic characteristics from other registers.

One challenge in the context of our research question is the measurement of stress. We cannot directly measure life-time stress. However, besides looking into overall mortality we can analyze two specific groups of medical diagnoses that have been related to stress during life in the literature: Cardiovascular diseases, specifically heart attacks and strokes, and smoking-related diseases, specifically lung cancer and chronic obstructive pulmonary disease (COPD). Stress from work-family conflicts is strongly predictive of smoking behavior (Nelson et al., 2012; Hurtado et al., 2016) and smoking behavior in turn is strongly correlated with a higher risk of dying from lung cancer and COPD. Low et al. (2010) survey the connection of stress to coronary heart diseases among women. They suggest that women that have to fulfill multiple roles in the family are more prone to suffer from these diseases. Ridker et al. (2000) have shown that among markers of inflammation, C-reactive protein (CRP) and interleukin-6 (IL-6) are strong predictors of cardiovascular diseases in older women. At the same time CRP and IL-6 are known to be elevated by chronic stress, such as care-giving (Kiecolt-

Glaser et al., 2003; Robles et al., 2005). Thus, in addition to overall mortality we can analyze specific causes of death that have been linked to stress during life in previous work.

Our paper is linked to past research on the interacting effects of fertility, working life, and maternal health. Using retrospective data from the Health and Retirement Study (HRS), Sabbath et al. (2015) categorize mothers along their past marriage, fertility and working histories into seven typical work-family profiles and investigate differences in mortality between these groups. Working single mothers experience the highest mortality rates in old age. Sabbath et al. (2015) extend the analysis with respect to mothers' job control and demand and find the highest mortality rate in the subgroup of mothers who became single later in life and had low job control. Van Hedel et al. (2016) investigate the association of work-family profiles with strokes, heart-diseases and smoking and compare estimates from the US and Europe. Again, single working motherhood is associated with higher likelihood of stress-related heart diseases, see also Berkman et al. (2015). While suggestive of a double burden effect, these studies are not able to control for selection of women into specific work-family profiles depending on their health. By using twins at first birth as a fertility shock we can overcome part of that problem.

Closest to our study are Cáceres-Delpiano and Simonsen (2012) and Kruk and Reinhold (2014), who study the relation of fertility and parental health outcomes. Using multiple births as an instrumental variable, Cáceres-Delpiano and Simonsen (2012) find that a higher number of children implies worse health for mothers aged 20 to 45 in the United States. Based on data from the Survey of Health Aging and Retirement in Europe, Kruk and Reinhold (2014) show that an increase in the number of children has a negative impact on mental health of older women but no effect on older men. The authors use twin births and sibling sex composition as instruments for the number of children.

Overall we find, that women who had twins at first birth are significantly more likely to die prematurely compared to mothers of singletons. At age 55 to 65 in

1990 Swedish twin mothers' have a 3.8 percentage point (13%) higher probability of dying within the next twenty years. Additionally we find that having twins at first birth significantly increases the likelihood of dying from heart attacks, strokes, lung cancer and COPD. The effects are significantly larger among highly educated mothers and those with above-median pension income at age 70. These results comes at a surprise, because highly educated mothers are a positively selected group with a significantly lower baseline probability of death. However, our findings are in line with the medical literature that studies the relation between stress, diseases and mortality. Specifically, our evidence concurs with an argument that the double burden of working and raising children increases life-time stress and takes its toll on mothers' health later in life.

This paper proceeds as follows. The administrative data set is introduced in section 3.2. Section 3.3 lays out our empirical strategy, while section 3.4 shows the results. We conclude in section 3.5.

3.2 Data

We use the Swedish multi-generation register, which links all individuals to their biological mother and father, even if they do not live in the same household or have died. It contains parental information for persons born in 1932 or later.[1] The multi-generational register has information on year and month of birth. Twins are identified as being born to the same mother in the same year and month as another sibling.

From the registry we identify 404,286 mothers that were 55-65 years old, alive and resident in Sweden in 1990. Of those, 2,684 mothers (0.66%) had twins at first birth. We exclude mothers with higher order births than twins. We can follow these mothers for twenty years from 1991 to 2010. From the death register we know if they died and we have information on the cause of death. We identify two

[1]For further information about this register see Ekbom, 2011.

specific groups of diseases that may be related to stress during life: Cardiovascular diseases and smoking-related diseases. The former comprise heart attacks and strokes and the latter lung cancer and COPD. We follow the strategy by Evans and Moore (2012) to classify the diagnoses into specific disease categories, see Appendix A for details.

Note that 8.6% of all mothers born between 1925 and 1935 are not observed in 1990 because they either died (75%) or moved abroad (25%) before 1990. In Appendix Table 3B.1 we investigate whether twin and non-twin mothers differ in the probability to be included in our study sample (column 1) or in the probability to die prematurely (column 2) and find no significant differences. Thus, while the sample as a whole may suffer from survival/migration bias our results are unlikely to be biased because twin and non-twin mothers are affected symmetrically.

Additionally we draw a rich set of socio-economic variables from the population register, for example, education and pension income. Table 3.1 describes our variables for the full sample and stratified by mothers' educational attainment. Education is defined in three categories. Primary schooling means that mothers completed compulsory education of nine years. Secondary schooling means that mothers had at least some years of secondary schooling. Tertiary schooling indicates that mothers experienced some tertiary schooling, i.e. some university education or even hold a PhD.

On average the mothers are 60 years old in 1990. They had their first child at age 24.5 and have on average 2.4 children. The majority of the mothers completed primary education (59%), about 30% of them hold a secondary and around 11% hold a tertiary degree. The age at first birth is on average three years higher in the highly educated group as compared to the low educated group, while the average number of children is about the same.

Overall, about 66% of the women between age 55 and 65 are still active on the labor market, i.e. receive a positive labor income in 1990. The fraction of working women varies considerably by education. While 89% of the women with a tertiary degree receive labor income, only 57% of mothers with a primary schooling degree receive income form work at those ages. About 29% of the women in our sample died between 1991 and 2010 with large variation by education. While about one third of the low educated mothers died in the 20 year time window we consider, the fraction is only 26% (19%) among the medium (highly) educated mothers. Thus, low educated mothers are roughly 1.7 times more likely to die over a 20 year period than highly educated mothers of the same cohorts. The prevalence of lung cancer and COPD, and heart attacks and strokes also decreases by education.

Pension income at age 70 follows the expected pattern; the mean pension income of mothers with tertiary schooling is just under 100,000 SEK and about 70% higher than the pension income of mothers with primary schooling or less.[2] The probability to receive above median pension income strongly increases with education.

[2] 100,000 SEK correspond to 10,752 EUR in 2002.

Table 3.1: Descriptive statistics by mothers' education

	Full sample	Primary schooling	Secondary schooling	Tertiary schooling
Age (1990)	60.03	60.34	59.74	59.24
	(3.16)	(3.13)	(3.16)	(3.06)
Age at first birth	24.56	23.92	24.81	27.13
	(4.67)	(4.55)	(4.62)	(4.43)
Number of children	2.40	2.45	2.32	2.36
	(1.21)	(1.30)	(1.11)	(1.00)
Twins at first birth (in %)	0.66	0.64	0.66	0.82
	(0.08)	(0.08)	(0.08)	(0.09)
Same-sex Twins	0.44	0.42	0.44	0.55
at first birth (in %)	(0.07)	(0.06)	(0.07)	(0.07)
Employed (1990 in %)	66.00	57.08	74.82	88.74
	(0.47)	(0.50)	(0.43)	(0.32)
Died between 1991 and 2010 (in %)	28.72	31.81	26.16	19.48
	(0.45)	(0.47)	(0.44)	(0.40)
Died from lung cancer	4.46	5.02	4.15	2.40
or COPD (in %)	(0.21)	(0.22)	(0.20)	(0.15)
Died from heart attack	13.61	15.70	11.85	7.53
or stroke (in %)	(0.34)	(0.36)	(0.32)	(0.26)
N	404,286	237,558	120,340	46,388
in %	100.00	58.76	29.77	11.47
Pension income at age 70	663	569	672	994
in 100 SEK	(360)	(320)	(322)	(385)
Pension income above median (in %)	50.00	40.19	52.89	79.92
	(0.50)	(0.49)	(0.50)	(0.40)
N	62,058	33,069	20,216	8,773

Note: For each variable the first line shows means with standard deviations below in parentheses. Primary schooling defined as education levels 1 and 2, Secondary schooling as level 3 and 4, and tertiary schooling as 5, 6 and 7.

3.3 Empirical strategy

We start by documenting all-cause and stress related mortality rates by the number of children a mother gave birth to in her life. While being purely descriptive, this analysis helps to understand the relationship between completed fertility and old-age mortality. We provide these estimates as previous studies have been inconclusive on the direction of this correlation (Hurt et al., 2006).

Our basic linear regression model takes the form

$$y_i = \alpha_0 + \sum_{k=2}^{K} \alpha_k \mathbb{1}\{\#kids_i = k\} + x_i' \alpha_x + \phi_i, \tag{3.1}$$

where y_i is the outcome variable. Depending on the specification, the outcome variables are indicators equal to one if the mother died from any cause between 1991 and 2010, if she died from a heart attack or stroke, or if she died from lung cancer or COPD between 1991 and 2010. $\mathbb{1}\{\#kids_i = k\}$ is an indicator equal to one if mother i gave birth to k children in her life, x_i are control variables and ϕ_i is an error term. We group mothers who gave birth to eight or more children into one category.

As controls we include dummy variables for seven different education levels.[3] We additionally insert dummies for mothers' birth cohorts, and a quadratic polynomial in age at first birth.[4]

We then turn to the causal effect of having twins at first birth on mother's

[3]Note that the levels of education we include in the regression model are finer than the three strata we use to condition our sample: 1=less then compulsory schooling of 9 years, 2=compulsory schooling of 9 years, 3=secondary schooling of at most 2 years, 4=secondary schooling of three years, 5=tertiary education of less than 3 years, 6=tertiary education of 3 years or more, 7=PhD.

[4]We showed in the previous chapter that the probability to give birth to twins follows an inverted U-shape across age at birth.

health later in life. We specify the following regression model:

$$Y_i = \beta_0 + \beta_1 \mathbb{1}\{\text{twins}_i = 1\} + x_i'\beta_x + \epsilon_i, \tag{3.2}$$

where $\mathbb{1}\{\text{twins}_i = 1\}$ is an indicator equal to one if mother i gave birth to twins at first birth.

While previous studies used the birth of twins as an instrumental variable (IV) for fertility, we study the reduced form effects of twinning and interpret them as being caused by a random event. There are two reasons for this. First, twin pregnancies and delivery are on average of a greater health risk to the mother than are singleton births. This might directly translate into higher old-age mortality rates (Rauh-Hain et al., 2009; Buhling et al., 2003). Second, twins might influence birth spacing which in turn might have a direct effect on mother's health. On the one hand, twins are themselves extremely close-spaced, and on the other hand, twins might influence the spacing of further children. Thus, even in the case where twins do not affect completed fertility in the long run, they may affect the dynamics of child bearing, and could change the interaction of child rearing and working (Heckman and Walker, 1990; Troske and Voicu, 2012). We therefore estimate a mixture of effects. Twinning affects the number of children the mother has over a specific period of time and it potentially directly influences mothers' health as argued above. Thus, we are not using the birth of twins as an IV for fertility.

One worry when comparing twin-mothers with non-twin mothers is that twinning might not be entirely random, as discussed in the previos chapter. This issue arises due to dizygotic (fraternal) twins, which become more likely with increasing age of the mother (Reddy et al., 2005; Fauser et al., 2005) and the use of in-vitro fertilization (IVF) treatments (Thurin et al., 2004). Monozygotic twins on the other hand are considered to be truly random (Tong and Short, 1998; MacGillivray et al., 1988). See Hall (2003) for an exposition on mono- and dizygotic twinning.

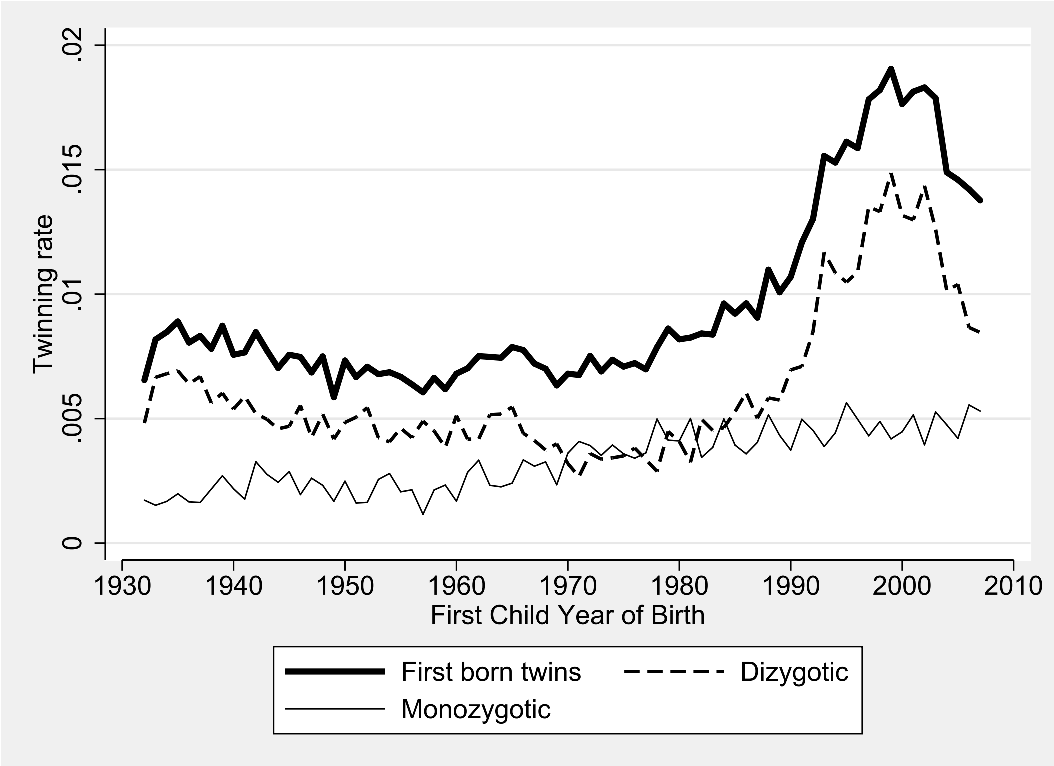
We have several strategies for dealing with these issues. First of all, as mentioned above, we control for age at first birth in all our analyses. Second, in order to investigate a possible selection bias stemming from dizygotic twins, we compare estimates between mothers of all twins and same-sex only twins (while also controlling for age at first birth). The reasoning is that since monozygotic twins necessarily have the same sex, their share must be higher among same-sex twins. The share of monozygotic twins among all twins is 33% and among the same-sex twins it is 50%.[5] See also Black et al. (2007) and Figlio et al. (2014), who find little differences in their estimates when using all or only same-sex twins. We follow this literature and present results for both measures.

IVF is less of a concern in our data, since more than 99% of the mothers gave birth to their first child between 1940 and 1970 when IVF treatment was not yet available. This is important, as the preference for IVF may be correlated with other health-related outcomes which we cannot control for. Figure 3.1 shows the twin rates across the first child's year of birth. The overall share of twins remains fairly constant between 1930 and 1980 but increases strongly thereafter. While the steady but mild rise in the twin rate after 1980 can be attributed to delayed child bearing, the steep increase in the twin rate since 1990 mainly follows the availability of IVF. Note that the fraction of monozygotic twins remains fairly stable over time.

In order to analyze the potential double burden effect we estimate the model in equation 3.2 on sub-samples defined by education and pension income. Education is an important predictor of labor force participation, working hours, and income. We can use education for stratifying the analysis since most of the mothers in our sample have completed their education before giving birth to their first child, i.e. the education level is less influenced by fertility than labor force participation itself.

[5] This can be see from Table 3.1. The share of all twin mothers in the sample is 0.66% and the share of same-sex twin mothers is 0.44%. Weinberg (1901)'s rule says that dizygotic twins are equally likely to be of same sex as of opposite sex. Under this rule, 0.22% of mothers gave birth to monozygotic twins.

Figure 3.1: Twin rate in Sweden (firstborn children) between 1930 and 2007.



Note: Statistics based on the Swedish register data. To compute the mono- and dizygotic twinning rates, we apply Weinberg (1901)'s rule.

We still checked if we have a sample selection bias by comparing age at first birth and schooling outcomes. As it turns out most mothers in Sweden have completed a schooling degree before their first birth in the cohorts considered. In our sample only 5.7% of mothers got their first child before age 19, which is the typical age of leaving secondary school. To explore this issue further we estimate an ordered logit model on education (with seven education levels) using twins at first birth, cohort fixed effects and a quadratic polynomial in age at first birth as explanatory variables. Column three of Appendix Table 3B.1 shows that neither all nor same-sex twins are significant predictors of education conditional on maternal age at first birth.

We have the following hypotheses with respect to education and fertility. While in principle the birth of twins lowers labor market activities of mothers of all educational levels, it is likely that highly educated mothers are more prone to pursue careers because of higher opportunity costs. We thus expect that, given the same unplanned fertility shock, higher educated mothers are more likely to experience a double burden of working life and child rearing. However, while highly educated mothers might be more likely to work and thus experience more stress, the Grossman (1972) model predicts that higher educated individuals are better at using medical care and might thus be more able to mitigate possible negative effects on their health. Thus, the overall potential double burden on the health of highly educated mothers compared to those of lower levels of education is not entirely clear. Additionally, low educated mothers may also experience a double burden effect as they might need to work more hours due to lower hourly wages compared to mothers with higher educational degrees and higher hourly wages.

While higher education only holds the *ex-ante* potential to a higher labor force attachment, pension income at age 70 is a proxy for life-time income and is thus an *ex-post* realization of the former. We use the earnings-related part of the pension income (*tilläggspension*) and do not include the basic pension (*folkpension*) in our pension income measure. Note that pension income in Sweden is independent of the partners' income. Again one could worry about selective sorting when splitting the sample in this way. Twins at first birth could directly affect pension income or survival and retirement until age 70. In Appendix Table 3B.1 (columns four and five) we show that there are no differences between (same sex) twin mothers and non-twin mothers with respect to whether the pension income at age 70 is missing or the size of the pension at age 70 conditional on covariates.

We expect a higher potential double burden effect on mothers with a higher pension income. Higher pension income reveals if an individual was active on the labor market during her life-time and thus was potentially affected by the double burden effect. We chose age 70 to avoid the selection problem that some individuals might still be working. The retirement age was 65 for the cohorts considered and only 20 individuals did not receive at least some pension income by age 70. Due to data restrictions, we can observe pension income at age 70 only for two cohorts, those aged 58 and 59 in 1990.[6] Thus, the sample for this analysis is considerably smaller. We stratify the sample at the median pension income in order to investigate the double burden effect.

3.4 Results

3.4.1 Completed fertility and old age mortality

As explained above, we first would like to investigate the correlation between the number of children and mothers' health later in life. In Panel A of table 3.2 we present OLS estimates of all-cause and stress related mortality rates of mothers aged 55+ using the model from equation 3.1. Mothers who gave birth to one child are the reference group. We find that overall mortality rates (column one) are significantly lower for mothers who have up to five children compared to those with only one child. Mortality rates are about equal between mothers with one and six children, but rise for mothers with more than six children. Thus, there exists a u-shaped relationship between the number of children and overall mortality. These patterns are in line with results by Grundy and Kravdal (2010) based on Norwegian register data.

In columns two and three we present cause of death specific results. For mothers with two to four children the number of children is associated with a lower likelihood of dying from lung cancer or COPD compared to mothers with one child (column two). In contrast to overall mortality, mortality rates from these diseases are not higher among mothers with more than five children, as compared to mothers with only one child. The coefficients for heart attack and strokes

[6] Pension income for individuals between age 65 and 74 is only available for the years 2001 and 2002.

(column three) again follow the pattern of all-cause mortality in the sense that mothers with up to four children show a lower likelihood of to dying within twenty years compared to mothers with one child only. Having more than four children is related to a significantly higher risk of dying from cardiovascular diseases.

Table 3.2: Mortality by number of children and twinning

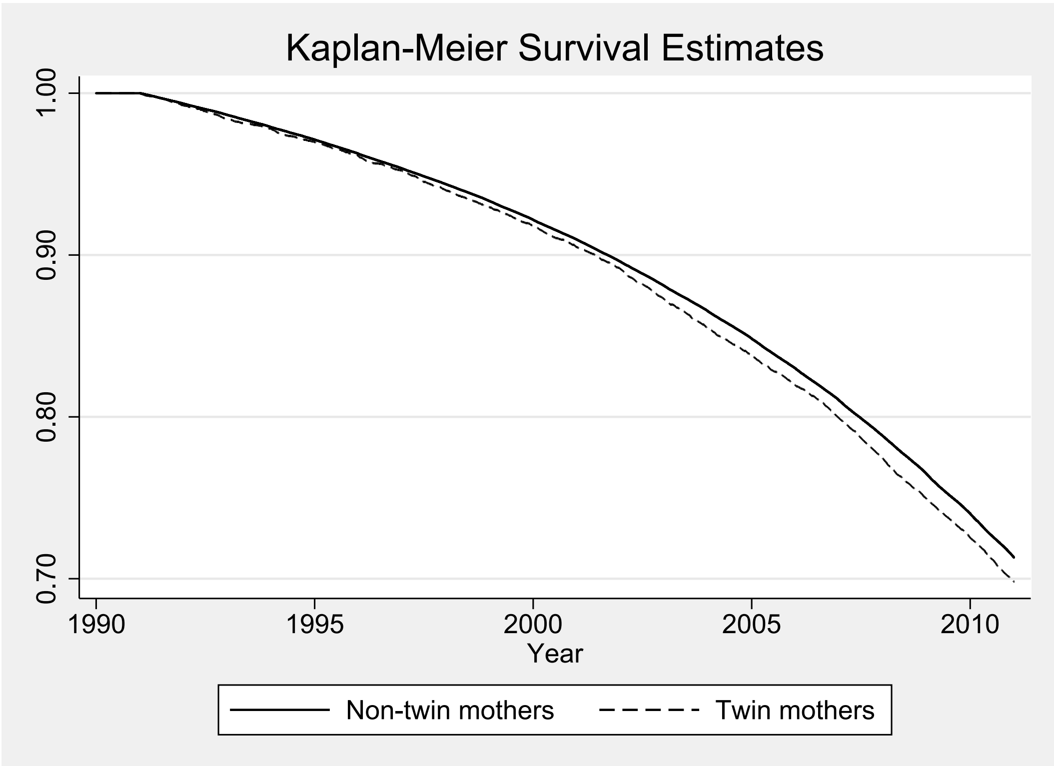
	(i) Died between 1991 and 2010	(ii) Lung cancer/ COPD	(iii) Heart attack/ stroke
Panel A			
Mothers with			
2 children	-0.042***	-0.010***	-0.022***
	(0.002)	(0.001)	(0.001)
3 children	-0.043***	-0.009***	-0.020***
	(0.002)	(0.001)	(0.002)
4 children	-0.030***	-0.004***	-0.009***
	(0.003)	(0.001)	(0.002)
5 children	-0.012***	0.000	0.007*
	(0.004)	(0.002)	(0.003)
6 children	0.005	0.002	0.020***
	(0.007)	(0.004)	(0.006)
7 children	0.032***	0.006	0.052***
	(0.011)	(0.006)	(0.009)
$\geq$ 8 children	0.041***	-0.000	0.058***
	(0.013)	(0.007)	(0.011)
Panel B			
Twins	0.037***	0.013***	0.018***
	(0.009)	(0.004)	(0.007)
Panel C			
Same-sex twins	0.038***	0.013**	0.020**
	(0.011)	(0.005)	(0.008)
Unconditional mean	0.287	0.044	0.136
Observations	404,286	404,286	404,286

Note: Table displays linear probability models controlling for education, cohort dummies and a quadratic polynomial in age at first birth. In panel A the reference group are mothers with one child, in panel B the reference group are mothers without twins at first birth and in panel C mothers without same-sex twins at first birth. Robust standard errors in parentheses below. ***,**,* indicate significance at the 1%, 5%, and 10% level, respectively.

3.4.2 Twins and old age mortality

We now turn to our main analysis. In Figure 3.2 we present Kaplan-Meier survival curves for mothers with and without twins. A clear gap in the survival probabilities emerges between the two groups over the 20-year period. Twin mothers are dying at a higher rate compared to their peers who only had one child at first birth. The gap becomes larger around the year 2000, i.e. when the women in our sample are on average 70 years old. We performed the same analysis using only the same-sex twin mothers and the pattern is even more pronounced.

Figure 3.2: Survival rates of mothers with and without twins 1990 to 2010



Panel B of table 3.2 contains estimates for the effect of having twins at first birth based on the regression model from equation 3.2. Having twins at first birth increases the probability of dying by 3.7 percentage points over a 20 year period. Related to a baseline probability of dying of 28.7% this means that twin mothers have a 13% higher mortality rate compared to mothers of singletons during the period of observation.

Looking into specific causes of death the pattern is confirmed. Twin mothers are 1.3 percentage points (or 20%) more likely to die of lung cancer or COPD compared to other mothers. Their likelihood of dying from a heart attack or stroke is 1.8 percentage points or 13% higher during the period of observation.

As a robustness check, Panel C contains estimates when using only same-sex twins as treatment, excluding potentially non-random opposite-sex twins from the treated group. The effects do not differ by much from the results presented in Panel B. This suggests that our results are not driven by non-random selection into twinning.

3.4.3 Results by educational level

In order to investigate the interaction of fertility and labor market attachment, we now split the sample by education and pension income. Table 3.3 displays our results stratified by maternal education. Panel A shows that the effect of having twins at first birth for mothers with at most primary schooling is slightly smaller compared to the effect estimated for the whole sample (see table 3.2 Panel B). However, the probability of dying from lung cancer, COPD or heart diseases is slightly higher among the mothers with a primary schooling degree compared to the overall effects. For mothers with a secondary school degree we find a similar effect of twins on overall mothers' mortality compared to the full sample (Panel B). However, there are no elevated levels of lung cancer and COPD or cardiovascular diseases for these mothers. The largest effect sizes in absolute and relative terms are experienced by mothers within the highest education group (Panel C). For twin compared to non-twin mothers all-cause mortality is increased by 8.4 percentage points or 43%, and death due to lung cancer and COPD is increased by 2.2 percentage points, which corresponds to an almost 100% increase. Death due to a heart attack or stroke is 4.1 percentage points or 55% higher. We find that effect sizes for the estimates based on all and same-sex twins are quite similar for all specifications.

Table 3.3: Results by education

	(i) Died between 1991 and 2010	(ii) Lung cancer/ COPD	(iii) Heart attack/ stroke
Panel A: Primary schooling			
Twins	0.028**	0.018***	0.021**
	(0.012)	(0.006)	(0.010)
Same-sex twins	0.027*	0.020**	0.023*
	(0.015)	(0.008)	(0.012)
Unconditional mean	0.318	0.050	0.157
Observations	237,558	237,558	237,558
Panel B: Secondary schooling			
Twins	0.032**	-0.003	0.001
	(0.016)	(0.007)	(0.011)
Same-sex twins	0.028	-0.005	-0.002
	(0.020)	(0.008)	(0.014)
Unconditional mean	0.262	0.042	0.119
Observations	120,340	120,340	120,340
Panel C: Tertiary schooling			
Twins	0.084***	0.022**	0.041**
	(0.023)	(0.011)	(0.017)
Same-sex twins	0.099***	0.020	0.055***
	(0.028)	(0.013)	(0.021)
Unconditional mean	0.195	0.024	0.075
Observations	46,388	46,388	46,388

Note: Each coefficient-standard error pair comes from a separate regression of a linear probability model controlling for education, cohort dummies and a quadratic polynomial in age at first birth. Robust standard errors in parentheses below. ***,**,* indicate significance at the 1%, 5%, and 10% level, respectively.

3.4.4 Results by pension income

Finally, we split the sample at the median of the pension income at age 70. As described in section 3.2, we can only use mothers aged 58 and 59 in 1990 for this exercise. Thus, the sample size drops considerably.

Results for overall mortality are shown in Table 3.4. Column one shows the effect of having twins at first birth on mortality. For mothers of twins the probability of dying is increased by 3.4 percentage points in the sample of 58 and 59 year old women. The effect size is quite close to the 3.7 percentage points estimated in the full sample (see Table 3.2). Comparing the effect of twinning for individuals below and above the median pension income demonstrates that the effect found in the combined sample is clearly driven by individuals with above median pension income (columns two and three). [7] Among mothers with a pension income above the median, having twins increases the probability of dying over a 20 year time period by 9 percentage points. Compared to a baseline probability of 15.2% this translates into an almost 60% higher mortality.

[7] As a robustness check we ran the same analysis for the pension income at age 69 for the cohorts age 57 and 58 in 1990 and the results are very similar.

Table 3.4: All-cause mortality by pension income.

Sample	(i) Pension at 70 observed	(ii) Low pension	(iii) High pension
Twins	0.034*	-0.016	0.090***
	(0.020)	(0.024)	(0.031)
Same-sex twins	0.077***	0.034	0.121***
	(0.026)	(0.034)	(0.039)
Unconditional mean	0.156	0.161	0.152
Observations	62,058	31,060	30,998

Note: Each coefficient-standard error pair comes from a single regression of a linear probability model controlling for education, cohort dummies and a quadratic polynomial in age at first birth. Low (high) pension sample are individuals with below (above) median pension income at age 70. Robust standard errors in parentheses below. ***,**,* indicate significance at the 1%, 5%, and 10% level, respectively.

3.5 Discussion and Conclusion

In summary, we find evidence of a u-shaped relationship between the number of children and mothers's mortality. This appears to be in line with healthy mothers getting more children. However, after a certain point more children take their toll on mothers' health either due to stress or because of the direct negative effects of giving birth to a large number of children on mothers' health. While this pattern cannot be interpreted causally, we also estimate the effect of having twins at first birth on mothers' mortality later in life. We find that having (same-sex) twins increases all-cause morality significantly for women older than 55. We also find substantial effects on cause-specific death rates. In particular, twin mothers have a higher probability of dying from lung cancer and COPD and heart attacks or strokes. The effects are stronger among women with higher educational degrees and higher pension income.

Our results fit into a recent line of epidemiological and sociological literature that tries to determine the adverse effects of work-family strain on women's later life health. The general theoretical considerations in that literature are the following. First, there are selection effects, i.e. women who are employed, married, and have children are healthier than their childless, unmarried and unemployed counterparts. Second, according to the role accumulation theory, combining family and work is beneficial for women's health. Third, multiple role theory states that combining work and family roles leads to stress with negative consequences for health.

Using twins at first birth as a shock to fertility we can overcome part of the endogeneity problem plaguing this literature. All-cause mortality as well as dying from lung-cancer and heart diseases are significantly elevated among mothers that give birth to twins. We take the higher probability of death due to lung cancer and COPD as well as the higher death rates due to heart attacks and strokes as indication that at least part of the effect is stress related. As argued in the introduction, the medical literature strongly associates these causes of death with stress from work-family conflicts, care-giving, and multiple role requirements of women. Thus, our results indicate that the additional burden on women due to getting two instead of one child at their first birth takes its toll on their health later in life.

In addition to overall higher levels of all-cause and stress-specific mortality among twin mothers we find particularly strong effects among mothers with tertiary education and above median pension income at age 70. These results come as a surprise, as these mothers are a socioeconomic advantaged group that due to high education, ability, income and savings should be more able to stay healthy and mitigate negative shocks (Smith, 1999). However, we take this as an indication of a double burden effect, because women with tertiary education have a higher likelihood of following their career despite having kids and higher pension income is an indicator that women worked more during their life-time. In other words, we use tertiary education as an *ex ante* predictor of higher labor market attachment and pension income as a proxy for *ex post* realized labor market activity. Higher all-cause and stress-related death rates among women with tertiary education and above median pension income point to the existence of a double burden from simultaneous child rearing and working on maternal health in old age. Women who have worked more over their life have higher mortality rates from the same fertility shock than others.

The particular mechanisms behind these findings deserve further research. However, it is important to note that we make these observations in Sweden, a country famous for its generous parental leave and child support policies that attempts to make labor market and fertility decisions as compatible as possible. What is more, our findings are particularly important in the light of the fact that women of younger generations are increasingly more likely to stay attached to the labor force and raise children at the same time (Goldin and Mitchell, 2017). Also, fathers' roles in supporting their families both financially and by taking a more active role in raising children are changing, too. Thus, more research is necessary in order to find adequate policies that buffer the negative consequences of a potential double burden from parents.

Appendix

3.A ICD codes of outcome variables

We follow the classification of diseases documented in table A-2 in the online appendix of Evans and Moore (2012), see table 3A.1 below.[8] The ICD-9 is applied in the years 1979 to 1998, ICD-10 from 1999 onwards.

Table 3A.1: ICD codes of causes of death and hospitalization.

Disease	ICD-9	ICD-10
Lung Cancer	162.2-162.5, 162.8-162.9	C34
Heart Attack	410	I21
Other Heart Disease	390-398, 402, 404, 411-429	I00-I09, I11, I13, I20, I22-I51
COPD	490-496	J40-J43, J44.0-J44.7, J44.9, J45-J48
Stroke	430-439	I60-I69

[8] Their web appendix can be found here `http://home.gwu.edu/~tim_moore/Evans_Moore_Restat_Appendix.pdf`

3.B Sample Selection

Table 3B.1: Selection into different samples.

Outcome	(i) Resident in 1990	(ii) Died in 1961– 1990	(iii) Education (ordered logit)	(iv) Pension 70 missing	(v) Pension 70 in 100 SEK
twins	0.005	-0.004	-0.014	0.000	-7.270
	(0.005)	(0.004)	(0.037)	(0.002)	(16.269)
same-sex twins	0.005	-0.005	0.046	0.000	-16.226
	(0.006)	(0.005)	(0.046)	(0.003)	(19.930)
Unconditional mean	0.914	0.065	2.224	0.835	616.547
Observations	444,197	444,197	404,286	404,286	66,742

In columns 1,2,4 and 5, each coefficient-standard error pair comes from a single regression of a linear probability model controlling for education, cohort dummies and a quadratic polynomial in age at first birth. Column 3 uses education in seven levels as outcome, is estimated using a ordered logit model and controls for cohort dummies and a quadratic polynomial in age at first birth. Robust standard errors in parentheses below. ***,**,* indicate significance at the 1%, 5%, and 10% level, respectively.

Chapter 4

Does Having Insurance Make Overconfident?*

Abstract: *Research on the role of behavioral biases in contract theory implicitly assumes that biases are stable. We show for the example of overconfidence that such biases may be malleable by the incentives provided even if incentives should not affect rational decision makers. Using a novel laboratory experimental design that allows to disentangle selection from incentive effects, we find that having insurance against losses in a real-effort task induces individuals to consistently overstate their performance relative to others. At the same time, we find no evidence that overconfidence plays a role in insurance choice.*

4.1 Introduction

Self-assessments and beliefs matter in decision making and contract design. Optimal decisions depend on correct self-assessments and well-calibrated beliefs. One important example is self-confidence in own ability and performance. In particular, overconfidence has been established as a relevant aspect in individual's economic

*This chapter is based on joint work with Joachim Winter and Martin Kocher.

behavior. For example, overconfidence has been found to predict excess market entry of entrepreneurs (Camerer and Lovallo, 1999), risky investment decisions of CEOs (Malmendier and Tate, 2005), and speculative trading (Scheinkman and Xiong, 2003). In the context of insurance, Sandroni and Squintani (2007) consider the Rothschild and Stiglitz (1976) model in the presence of overconfident individuals. They find that if the share of overconfident types in the population is large enough, compulsory insurance is not Pareto-optimal anymore. It follows that overconfidence as a behavioral inclination has important implications for contract design in many settings (see for example Sautmann, 2011, De la Rosa, 2011 and Santos-Pinto, 2008).

Overconfidence and imperfect self-confidence calibration relate to many effects observed in human decision making. Our focus here is on overplacement, which is related to the better-than-average effect. However, a general interpretation of the literature on self-confidence is that over- or underconfidence are comparably stable traits, at least within a certain decision environment. That is, one can be overconfident when driving and underconfident with math tasks, but overconfidence when driving should not be affected by the color of the car. This paper provides evidence for self-confidence to be malleable in a setting that has relevant implications. We show in a laboratory experiment that confidence in one's own performance depends on whether people acquire insurance or not. While insurance in our setup partially covers potential losses from bad performance in a real-effort task, it should be unrelated to performance and to the overconfidence elicitation for rational decision makers. At the same time, we find no evidence for more confident individuals choosing more or less insurance in the first place.

More specifically, we implement an experimental design that allows us to cleanly disentangle effects from the incentives provided by the insurance contract from effects coming from selection into the contract. In the insurance context, the former is known as moral hazard and the latter as adverse selection. Before attempting the real-effort task, individuals are given the choice to buy an insurance contract. Conditional on this choice, the actual insurance status is randomized, i.e. whether one obtains insurance or not is based on a random draw, and individuals are informed about their insurance status. Our design is similar to the one used in a credit market field experiment by Karlan and Zinman (2009). Their idea is to attract borrowers with an advertised interest rate and, conditional on showing up in the lenders office, to randomize the actual interest rate. However, Karlan and Zinman (2009) are not able to impose an interest rate that is higher than the one advertised, as borrowers could simply walk out of the experiment. In a laboratory experiment, by design there is no attrition. This allows us to assess whether the effect of insurance on overconfidence only comes from feeling (un-)lucky when actually (not) receiving it - remember, insurance status is based on a random draw - or whether there is another mechanism that is able to explain the effect. A related design is used by Bó et al. (2010), who let individuals vote on a policy that allows punishment for defection in a prisoners dilemma, but then randomize the actual implementation of the policy (see also Sutter et al., 2010).

Our real-effort task involves the forecasting of numbers with the help of two cue values (Brown, 1998; Vandegrift and Brown, 2003; So et al., 2017). This task fulfills two requirements for our purpose of creating a realistic insurance setting. First, the ability for forecasting, which might in the present case be related to math skills, varies sufficiently in the sample to create different levels of confidence. Second, the participant's effort can influence the precision of their forecasts and thus their belief in their own performance. Schram and Sonnemans (2011) also consider insurance choice by varying various parameters such as the number of available contracts. However, in their setting, losses occur without a subject's influence, which may not be realistic for some insurance contracts such as car insurance. Previous experiments studied insurance choice with exogenous loss in various settings, see for example Ganderton et al. (2000) and Laury et al. (2009). Our design naturally exhibits features of insurance markets outside the laboratory such as adverse selection and moral hazard. Confidence is measured as an individual's self-assessed performance, relative to others in the real-effort task, stated as a rank within the experimental session. The elicitation is incentivized by rewarding accuracy. The form of overconfidence that we measure is termed overplacement (see Moore and Healy, 2008). On average, our subjects are slightly underconfident. This is in line with experiments by Clark and Friesen (2009) and Murad et al. (2016), who argue that the use of real-effort tasks and incentivized confidence elicitation lead to a lack of overconfidence which is generally observed in "better-than-average" predictions. Moore and Cain (2007) and Hoelzl and Rustichini (2005) find that subjects are underconfident in tasks that are perceived as difficult and where performance is low in absolute terms, which is in line with our setup.

Our contribution is threefold. First, we show that self-confidence can be affected strongly by actually irrelevant aspects of contractual design. While in its generality, this result is probably not too surprising, its impact on our insurance application bears relevant implications - just imagine that drivers become relatively more overconfident after being insured. While contract design has started to take behavioral biases into account (Kőszegi, 2014), we are not aware of any existing model that would be consistent with our main finding. Second, we experimentally study assumptions made on the selection mechanism into contracts based on presumably stable personality traits such as self-confidence calibration (see for example Sandroni and Squintani, 2007, 2013). This paper thus speaks to a broader literature that studies sorting into contracts based on behavioral biases and preferences (Larkin and Leider, 2012; Dohmen and Falk, 2011). Finally, we add experimental evidence to decision making in a behavioral insurance context in which own effort instead of a random device determines losses (Browne et al., 2015). We believe that such a setup adds to the external validity of our results for certain insurance classes.

4.2 Experimental design

We start by describing the general procedure in our experiment, the real effort task and then the insurance decision. Monetary payoff was based on points, converted to euros at a fixed and pre-announced exchange rate. Participants received an endowment of 100 points, equal to EUR 10. The show-up fee for participants was EUR 4. The experiment was computerized with the help of z-tree (Fischbacher, 2007), and participants were invited with the organizational software ORSEE (Greiner, 2015).

4.2.1 Experimental procedure

All steps in the experimental setup were known in advance and common knowledge among participants. However, we did not announce that we would elicit self-confidence after the real-effort task and insurance decision. The experiment consisted of three parts, and participants were aware of the existence of the three parts from the start of the experiment. They did, however, not know anything about the content of the following part until the end of the previous part. In the following, we just report results from the first part.[1] The experimental procedure for the relevant stages is illustrated in Figure 4.1, along with the variables generated at each stage. We explain the details for each stage below and in the subsequent sections.

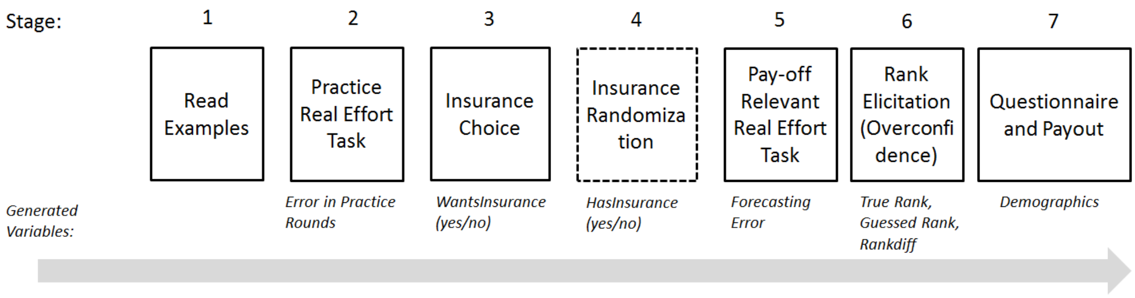
In the first stage, subjects received a sheet of paper with ten examples of solutions in the real-effort task. The real-effort task was a forecasting task, and participants saw realized values of Y, W_1 and W_2, which could be studied for five minutes, on the example sheet. A pen was provided, and participants were allowed to take notes, which was done frequently. The second stage consisted of five practice rounds (five forecasts) with feedback on individual performance. These

[1] The second part consisted of a set of lottery decisions; the third part was a short survey on relevant experience with insurance. Experimental instructions for the first part are provided in Appendix 4.C, and screenshots of steps 2 to 6 of the procedure can be found in Appendix 4.B.

practice rounds were not incentivized, but there was an implicit incentive in the form of a potential information gain regarding one's own ability in this task. In the third stage, individuals had to decide whether they wanted to buy the insurance for the upcoming payoff-relevant rounds or not. An on-screen calculator could be used at this point. The fourth stage randomized actual insurance receipt, and the choice made in stage 3 was realized with 70% chance. Thus, if a subject did not want to buy insurance, there was still a 30% chance that she got the insurance and that she had to pay the premium. Conversely, there was an equally large chance to not receive insurance, although the subject wanted to buy it. This creates a 2 by 2 matrix of possible outcomes shown in table 4.1. The probability of 70% was chosen trading-off incentive-compatibility and statistical power. A message informed participants about the realized insurance status. The message stayed on the screen throughout the following ten payoff-relevant rounds of the real-effort task in stage 5.

After the ten rounds of the real-effort task were completed, we elicited self-confidence in stage 6. Remember that this stage was not announced in the instructions. Individuals were asked to think about their average performance in the previous ten rounds and should indicate which rank they think that they hold in their respective session. The person with the lowest average forecasting error would take the first rank, the one with the second-lowest the second rank, and so on. At this point, subjects had not received any feedback on their or other participants' performance. Guessing the rank correctly earned 10 additional points, and a deviation of plus or minus one from the realized rank earned 5 additional points. We chose to measure confidence in performance after the task, instead of before the task, in order to avoid hedging behavior and possible priming effects. Asking individuals about their relative performance to others before the task could give the wrong impression of a competitive environment, which we neither consider in this paper, nor is it common in an insurance context. We are well aware of the fact that linear incentives when eliciting beliefs have their limitations (see, Gächter

Figure 4.1: Experimental procedure and definition of variables.



and Renner, 2010; Trautmann and Kuilen, 2015), but for our case it seems a good compromise between validity and straightforward implementation. Between stages 6 and 7, the second and third parts of the experiment took place. In stage 7, one of the ten real-effort task rounds was randomly drawn by the computer, and subjects were informed about their performance and earnings in this round. They also learned how much they earned from the ranking guess. At the end of the experiment, individuals answered a standard demographic questionnaire and were paid out in private.

Table 4.1: Sample distribution

Insurance status		actual yes	actual no	Total
choice	yes	68	41	109
		41%	25%	
	no	13	45	58
		8%	27%	
	Total	81	86	167

4.2.2 More information on the real-effort task

We used the forecasting task by Brown (1998), Vandegrift and Brown (2003), and So et al. (2017). Participants are asked to enter the price Y of a fictitious stock whose price they had to predict from two cue values W_1 and W_2. The true relationship of Y and the two cues was given by

$$Y = 50 + 0.3W_1 + 0.7W_2 + e,$$

where $W_1, W_2 \sim U(0, 250)$ and $e \sim N(0, 5)$. Y was rounded to the nearest integer. Individuals knew that there was a potential constant, but did neither know that the function was linear, that the weights added to one, nor that there was a random error term e. During the task, individuals where shown W_1 and W_2 on the screen and had 60 seconds every round to enter their forecast $\hat{Y}$ into a box and click OK (see figure 4B.5 in the Appendix). The remaining time was always displayed on screen. There were no incentives for speed, but after 60 seconds without any input the program would skip to the next round, automatically creating a no-input. We introduced a penalty to avoid this, and the details are described in the next section. From the forecasting input we derived the error in each forecast, which is given by the absolute difference between the true and the predicted value of Y:

$$error = |Y - \hat{Y}|$$

4.2.3 Insurance

Based on a pilot of the real-effort task, we set the insurance premium to 22.5 points, with a coverage 65%. Remember that only one round was payoff-relevant, i.e. the insurance was valid for all rounds. Earnings from the task are

$$earnings_{no} = 100 - error$$

for individuals that did not get the insurance and

$$earnings_{in} = 100 - error \times (1 - 0.65) - 22.5$$

for those that did. Thus, insurance covered 65% of the loss from the absolute difference between the true and the predicted value of Y. Notice that we capped losses at the zero earnings boundary. As a consequence, there were no losses from this part of the experiment unless a participant had not entered any forecast at all for the randomly chosen round and was insured. In that case, the participant would have to pay the insurance premium of 22.5 points from her show up fee. This happened only once.

4.2.4 Experimental participants

We conducted seven sessions in November 2015 in the MELESSA laboratory at the University of Munich. In total, 167 subjects participated and earned on average EUR 12.50 in a bit more than one hour per session. Participants were mainly students from various fields of study, with 33% from economics or business, 18% from life sciences or engineering and 13% from humanities. Almost 60% of participants were female, and age ranged from 18 to 43, with an average of 22.

4.3 Results

4.3.1 Descriptive results on overconfidence and insurance choice

We first look at a set of descriptive results. Our variable of interest is *rankdiff*, the difference between the individual's actual and guessed ranks as entered in stage 6 of the experiment:

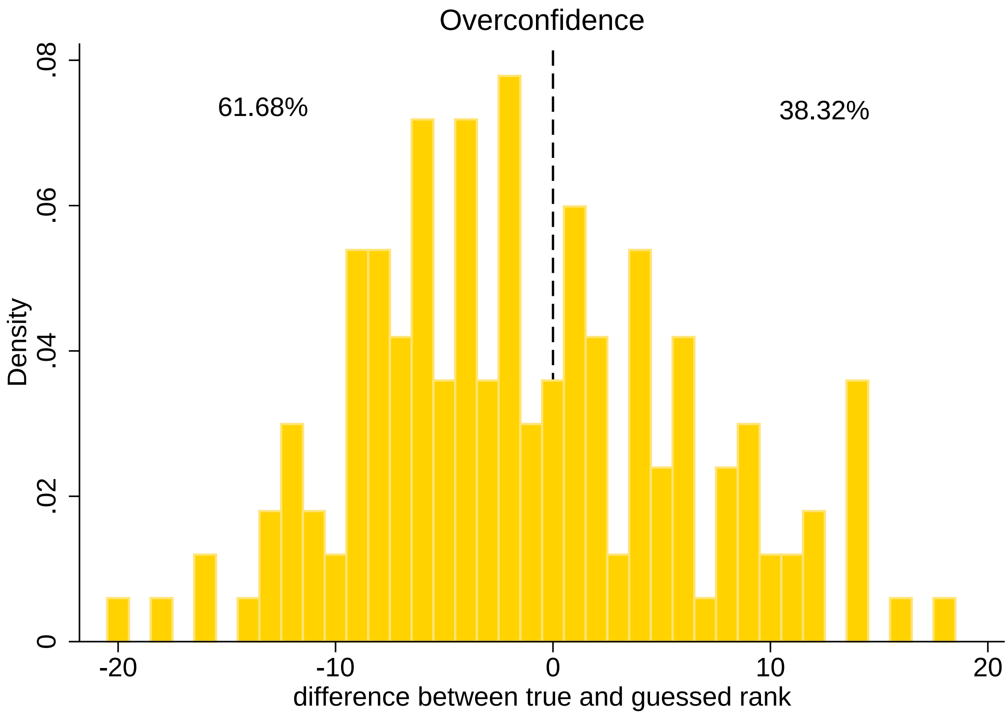
$$Rankdiff = TrueRank - GuessedRank.$$

A positive value indicates overconfidence, where higher values imply stronger overconfidence. A similar variable has been applied by Sautmann (2011), who uses the difference between predicted and actual scores in trivia quizzes as her measure for overconfidence. The mean of *rankdiff* in our study is -1.37 (which is significantly different from zero at the 5% level), indicating slight underconfidence, on average. The distribution of *rankdiff* is shown in figure 4.2. The average underconfidence result is in line with Hoelzl and Rustichini (2005) and their task-specific explanation. However, there exists considerable variation of self-confidence in our sample on the individual level and when comparing treatments. An alternative measure is a simple indicator variable for overconfidence. It takes on the value one if *rankdiff* is larger than zero, and the value zero otherwise. The entire sample has a share of 38.32% overconfident individuals according to this measure.

Remember that we can distinguish between four insurance outcomes, indicated by the variables *HasInsurance* and *WantsInsurance*. The variable *HasInsurance* describes the true insurance status of an individual in the real-effort task, and it is randomized. The variable *WantsInsurance* describes the individual's initial choice for or against insurance, and it is endogenous in the sense that it may correlate with any observed or unobserved individual characteristics such as gender, age and risk attitude. Conditional on insurance choice (=*WantsInsurance*), *HasInsurance* identifies the incentive effects of the insurance contract. Conditional on actual insurance status (=*HasInsurance*), *WantsInsurance* identifies selection effects, i.e. differences between individuals who wanted insurance and those who did not.

Table 4.2 displays means and standard deviations by insurance outcome. Table 4A.1 in the Appendix contains p-values of t-tests within every cell of table 4.2 for the hypothesis that the mean of *rankdiff* is significantly different from zero.

Figure 4.2: Distribution of variable rankdiff.



In addition, table 4A.2 displays p-values of pairwise, two-sided Wilcoxon-Mann-Whitney tests for differences in *rankdiff* between all experimental groups. We observe strong and highly significant underconfidence without insurance. There is, however, also significant underconfidence for those who did not want insurance, when we pool observations for those who ended up with insurance and those who did not.

Two-third (109 out of 167) of individuals wanted to buy the insurance. We can investigate which individual characteristics predicted insurance choice. Table 4.3 shows mean values of these variables by insurance choice status and in the full sample. Individuals who made larger errors in the practice rounds were more likely to want insurance, which is in line with standard predictions of adverse selection models. *Insurance pays off* is a dummy equal to one if the forecasting error in a practice round was larger then 22.5/0.65=34.62, which is the break-even point

Table 4.2: Mean and standard deviation of rankdiff

	Wants Insurance=1	Wants Insurance=0	Total
Has Insurance=1	0.088	-0.46	-2.01
	(7.39)	(6.00)	(7.67)
Has Insurance=0	-2.88	-2.46	-1.03
	(6.99)	(7.96)	(7.41)
Total	-2.66	0.00	-1.37
	(7.56)	(7.23)	(7.50)

(error) of the insurance for a fully rational, risk-neutral decision maker. There is a large difference (20%-points) between those who wanted insurance and those who did not. However, buying insurance would still have paid off in 40% of rounds for those that did not want to buy insurance. Females more frequently wanted insurance than males and so did younger individuals.

Table 4.3: Insurance choice

	Did not want insurance	Wanted insurance	Total
Error in practice rounds	41.52	57.81***	52.15
Insurance pays off	0.40	0.60***	0.53
Female	0.36	0.67***	0.56
Age	23.33	21.42***	22.08

Insurance pays off is a dummy equal to one if the forecasting error in a practice round was larger then 22.5/0.65=34.62, which is the break-even point (error) of the insurance for a fully rational risk-neutral decision maker. Stars indicate mean differences significant at 1% (***), 5% (**), and 10% (*) level. Standard errors clustered at individual level in rows 1 and 2.

4.3.2 Regression analysis

We now turn to the effect of insurance on overconfidence and selection into insurance based on overconfidence by using parametric models. All regressions in table 4.4 use OLS estimations and include session fixed effects.[2] We start with performance in the real effort task in the first column. We find that having the insurance increases the absolute forecasting error by 4 points (or 0.15 standard deviations). The same difference is found between individuals who wanted and did not want insurance. The first effect is moral hazard and the second adverse selection, two classic elements in insurance markets (Shavell, 1979; Rothschild and Stiglitz, 1976). Column two shows the direct consequence of a lower performance in the task: both incentive and selection effects lead to a higher (i.e. worse) ranking within a session. Column three concerns the rank that individuals guessed they are taking. Individuals who ultimately got the insurance do no rank themselves worse or better than those who did not. In contrast, the pure selection effect in guessed ranks equals the one in true ranks. It follows in column four that insurance increases the difference between individual's guessed and actual rank by 2.7 ranks. Conditional on actual receipt, there exists no significant difference between those subjects that wanted and did not want the insurance. This is in contrast to Sandroni and Squintani (2007), who assume that overconfident individuals are less likely to buy insurance, because they perceive their risk to be lower than is actually the case. We find that, on average, individuals anticipate their performance in the task based on their skill level and adjust their rank accordingly, but independent of the actual insurance status.

In the following we investigate if other biases specific to the experimental environment drive our results. One explanation could be that not getting the insurance despite wanting it leads to what is called "choking", a sudden decline of concentration and performance when individuals feel under pressure (Baumeister, 1984).

[2]Ordered logit (for rank outcomes) and logit (for the overconfident dummy) models yield very similar results. The results are available on request.

This could lead to a severe underestimation of own performance, independent of its true level. Conversely, individuals receiving the insurance might feel lucky and thus rank themselves better than they actually are. These two confounding factors imply that the effect of the insurance on overconfidence should be larger among those individuals who also wanted it. In our 2 by 2 design, we can test for this possibility. Column five shows that the interaction term between wanting and actually receiving the insurance is positive, but far from significant. The main effect of the insurance is not significant anymore, but the point estimate is similar to that in the columns before.[3] Column six includes gender and age as explanatory variables to check if these explain the non-significant selection effect. Although the coefficient turns positive, it is not statistically significant and only one-third of the insurance effect. Columns seven and eight replicate columns four and six with a dummy equal to one if *Rankdiff* is positive as outcome variable and we get qualitatively similar results. The occurrence of overconfidence in ranking is increased by one-quarter under the insurance contract.

4.4 Discussion

One major concern when trying to elicit self-assessment biases is to detect what Benoît and Dubra (2011) call apparent overconfidence. If individuals are Bayesian updaters and receive only a limited number of noisy signals on their performance, they might rationally rank themselves better than others, while this is interpreted as overconfidence by the researcher. This is less of a concern in our experiment, as individuals do not receive any signal on their (or others') performance in the payoff-relevant rounds. Their ranking should therefore solely be based on the perceived difficulty of the task over the ten rounds and an idiosyncratic component, which

[3]This could also be due to lack of power, as the main coefficient of *HasInsurance* now refers to the insurance effect in the group that did not want the insurance and this group comprises only one-third of the sample. The insurance effect in the group that wanted the insurance is still significant at the 10% level.

Table 4.4: Insurance and overconfidence

	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)
Outcome:	Error	True rank	Guessed rank	Rankdiff			$\mathbb{1}\{\text{Rankdiff} > 0\}$	
HasInsurance	4.088**	2.311**	-0.649	2.960**	2.443	3.157**	0.240***	0.251***
	(1.729)	(1.147)	(0.872)	(1.235)	(2.137)	(1.254)	(0.082)	(0.083)
WantsInsurance	4.032***	3.081***	3.303***	-0.222	-0.473	0.925	-0.016	0.042
	(1.544)	(1.177)	(0.893)	(1.262)	(1.710)	(1.400)	(0.084)	(0.091)
Has × Wants Insurance					0.729			
					(2.709)			
Female						-1.651		-0.016
						(1.329)		(0.080)
Age						0.391**		0.031***
						(0.171)		(0.010)
Constant	18.171***	9.368***	11.341***	-1.974	-1.943	-11.268**	0.296**	-0.475*
	(2.407)	(1.730)	(1.118)	(2.174)	(2.187)	(4.793)	(0.114)	(0.263)
Session f.e.	yes	yes	yes	yes	yes	yes	yes	yes
N	1,670	167	167	167	167	167	167	167
Adj. R-squared	0.017	0.056	0.053	0.000	-0.006	0.028	0.032	0.074

Rankdiff is the difference between the true and guessed rank of performance in the task. Individuals were incentivized to guess their rank among all participants in their session with respect to their average performance in the 10 payoff-relevant rounds of the forecasting task. No feedback on performance was provided. Robust or clustered (column one) standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

on average is the same between those that get and do not get the insurance, conditional on choice. Furthermore, Merkle and Weber (2011) demonstrate that the extent to which apparent overconfidence poses a problem in the laboratory is limited.

Another concern may be an insurance-induced change in a potential hedging motive when confidence levels are elicited. Since insurance reduces the downside risk in the real-effort task, the hedging motive in the confidence elicitation part loses importance. As a result, insured individuals could understate their performance less strongly than non-insured. However, this would imply that the insured place themselves at better ranks than the non-insured, which is not the case, as can easily be seen in column three of table 4.4. Another change in placement behavior arises if participants anticipate the lower performance of others, potentially induced by having insurance. Knowing that others will perform worse, they can place themselves better in the confidence elicitation. However, such higher order thinking applies to both treatment groups and should therefore be averaged out.

4.5 Conclusion

In this paper, we conducted a laboratory experiment in which losses from a real effort task could be reduced by purchasing an insurance. Conditional on choice, actual insurance receipt was randomized, allowing us to disentangle selection from incentive effects. Overconfidence is measured as the difference between an individual's true and self-assessed performance rank. While the previous literature is concerned about selection, we are the first to demonstrate that normatively irrelevant incentives can change overconfidence ex-post. Moreover, we find no evidence for selection into insurance based on overconfidence.

Why does the insurance make individuals relatively overconfident in their performance? One possible explanation from our regression analysis is that individuals do not anticipate the moral hazard that is introduced by the insurance. Subjects do however anticipate their skill level and adjust their rank estimate accordingly. Put differently, the effect of the insurance is not reflected in an adjusted ranking, while the selection effect is. Another explanation involves the perception of the difficulty of the task. Under insurance, the task could appear easier, although it is actually only the loss from the task which is lowered. As a consequence, underplacement is reduced. One can imagine alternative psychological explanations: for instance, insurance could let individuals focus more strongly on potential gains and thus, the expected performance could appear more gloomy.

Our results have implications for insurance markets. Take car insurance as an example. Outside the laboratory it is next to impossible to distinguish between potential moral hazard effects and potential overconfidence effects. If both are present, the optimal policy of the insurer should take both into account. Remedies against moral hazard would not be enough to minimize unwanted behavioral tendencies, when we assume that biased self-assessment has negative consequences on driving. The experiment in this paper also has its limitations. For reasons explained above we do not have measures of overconfidence before randomization of the insurance. Further, we have no information on whether the induced overconfidence translates to other tasks and situations without insurance or on whether it is persistent or not. Ultimately answering this puzzle will require further research on why individuals become overconfident in the first place.

Appendix

4.A Additional Figures and Tables

Figure 4A.1: Distribution of forecasting errors in practice and payout-relevant rounds.

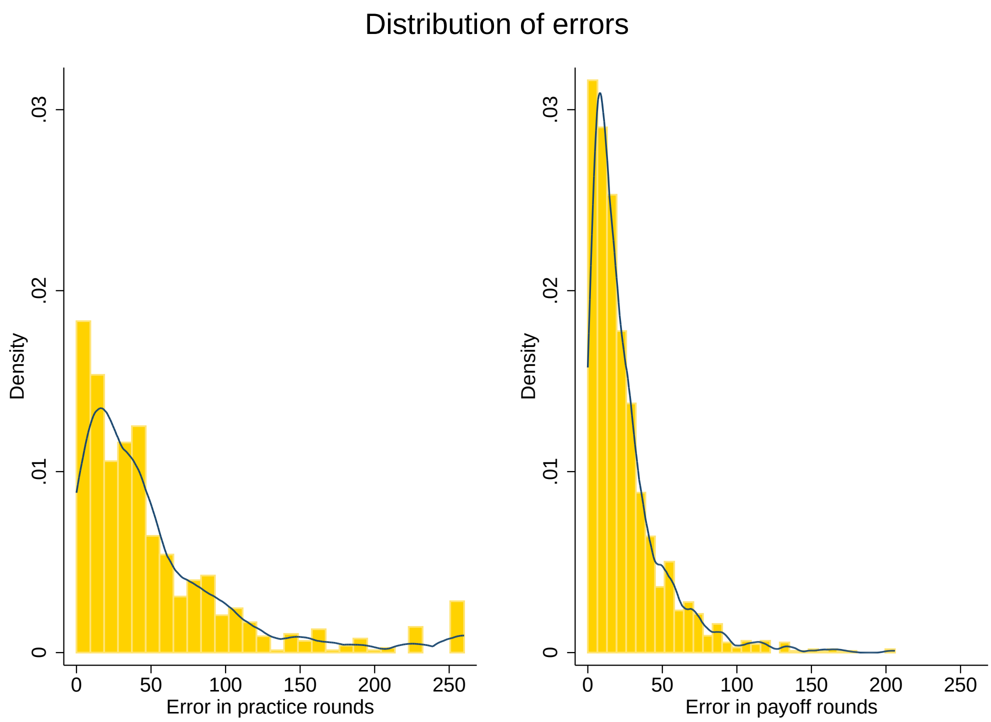


Table 4A.1: P-values for zero mean t-test of rankdiff

	Wants Insurance=1	Wants Insurance=0	Total
Has Insurance=1	0.922	0.794	1.000
Has Insurance=0	0.013	0.046	0.002
Total	0.151	0.050	0.019

Notes: Table shows p-value from t-test with the Null hypothesis that the mean of rankdiff equals zero within the respective cell.

Table 4A.2: P-values from Wilcoxon-Mann-Whitney test of pairwise difference in rankdiff

Group 1	Group 2		p-value
has=1	has=0		0.021
wants=1	wants=0		0.445
has=1	has=0	| wants=1	0.051
has=1	has=0	| wants=0	0.287
wants=1	wants=0	| has==1	0.862
wants=1	wants=0	| has==0	0.839

Notes: Table shows p-value from Wilcoxon-Mann-Whitney test of a difference in rankdiff between experimental groups.

4.B On-screen instructions

Figure 4B.1: Stage 2a: The real effort task in practice rounds.

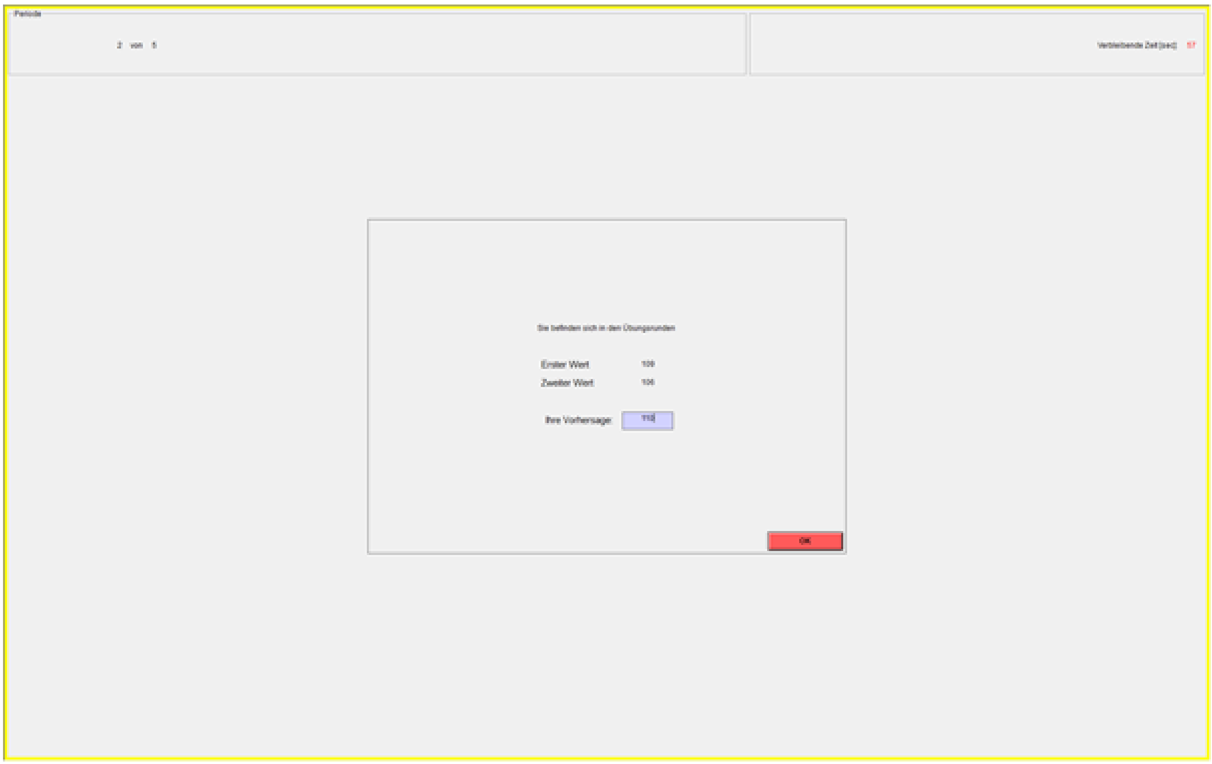


Figure 4B.2: Stage 2b: Feedback to the real effort task in practice.

Figure 4B.3: Stage 3: Decisions whether to buy the insurance.

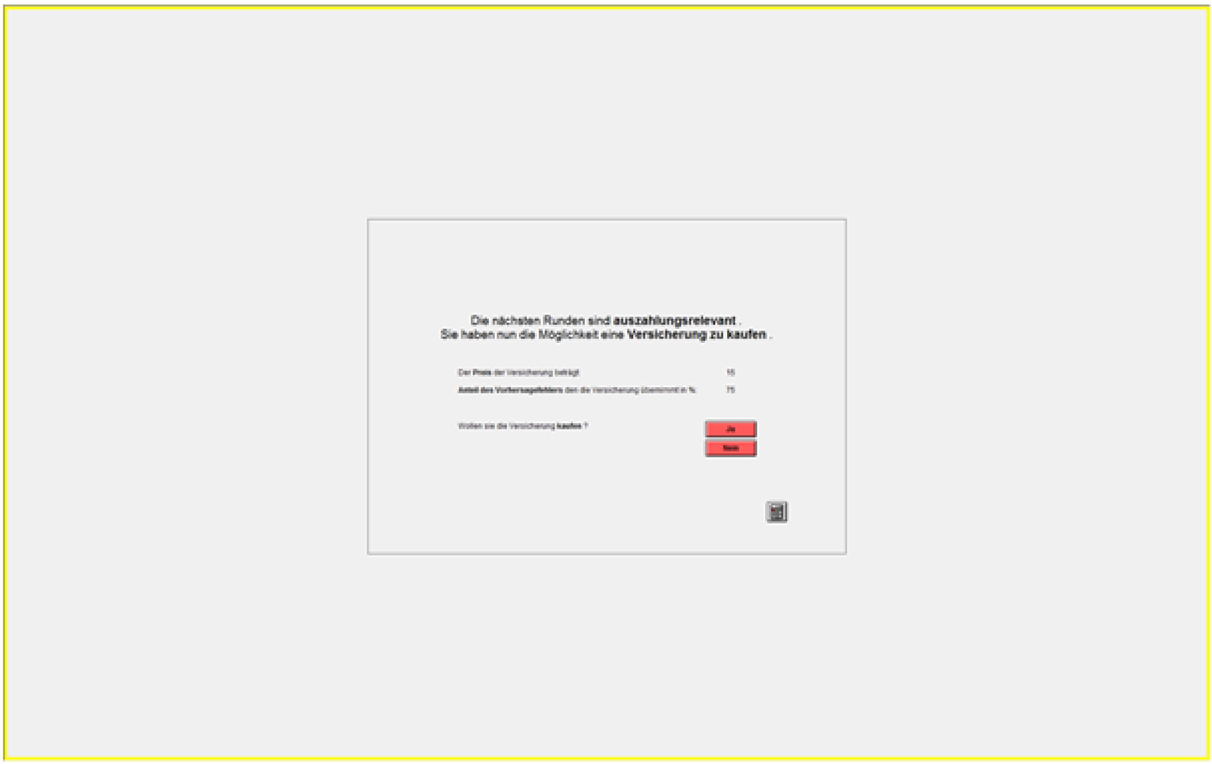


Figure 4B.4: Stage 4: Message on realized insurance status.

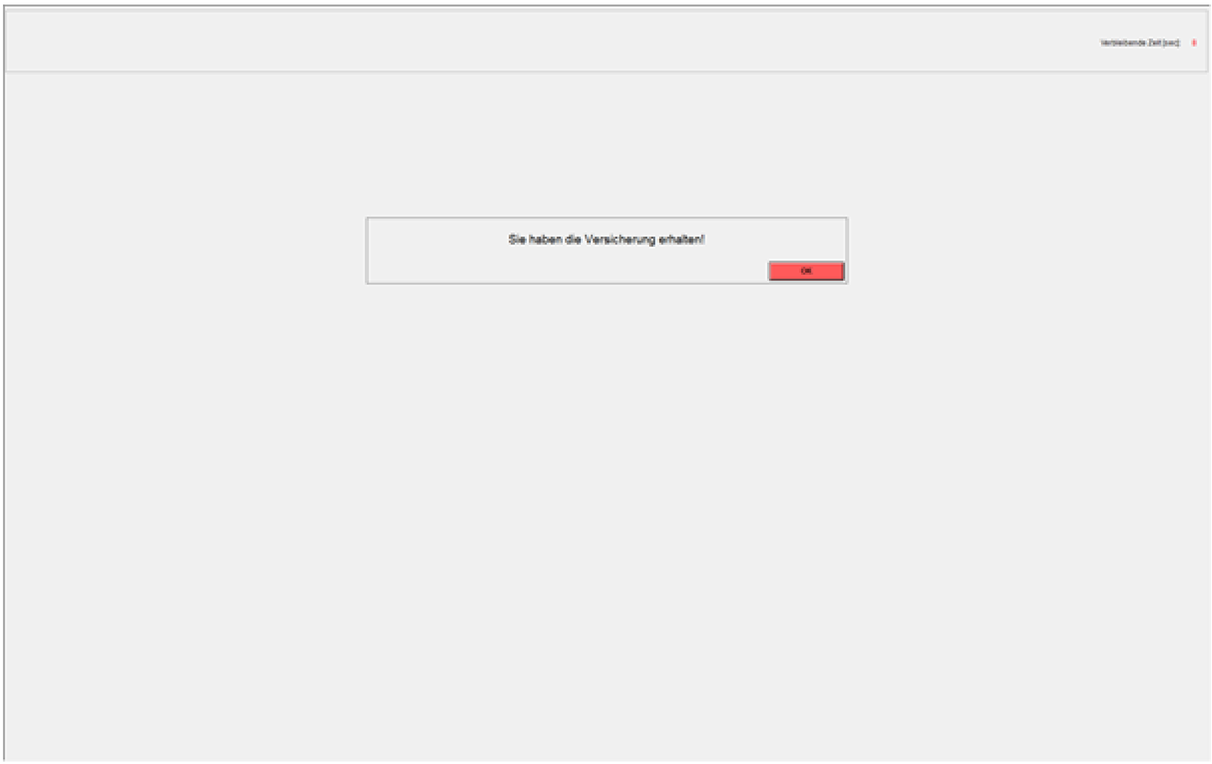


Figure 4B.5: Stage 5: The real effort task in payoff-relevant rounds.

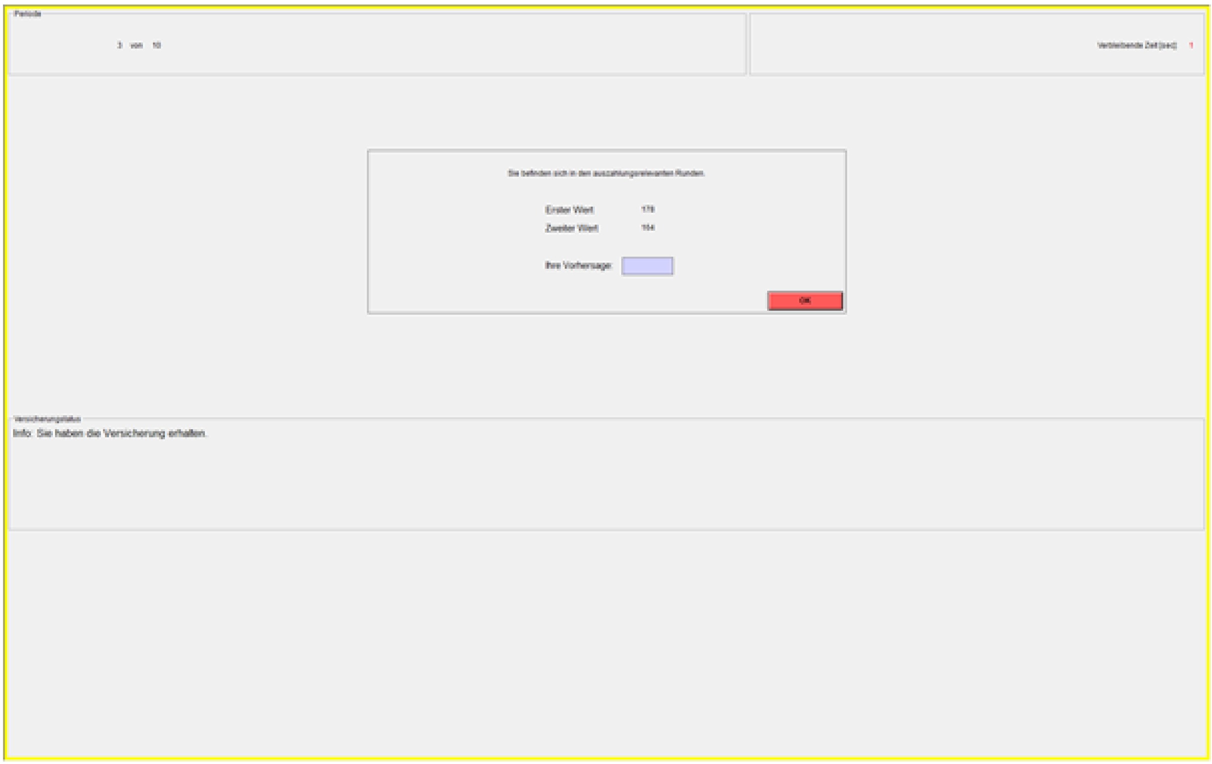


Figure 4B.6: Stage 6: Ranking of own performance within session.

4.C Experimental instructions

Instructions are translated from German. Instructions were identical for all participants. Instructions from the second part of the experiment are not shown here.

Welcome to the experiment and thank you for your participation!

Please stop talking with the other participants now

General procedures

In this experiment we study economic decision making. You can earn money by participating. The money you earn will be paid to you after the experiment privately and in cash.
The experiment takes about 1 hour and consists of three parts. At the beginning of each part you will receive detailed instructions. If you have any questions about the instructions or during the experiment, please raise your hand. An instructor will then come to you and answer your questions privately.

Payment

You profit will be denoted in points, where 10 points = EUR 1. In part I and II you will have to solve multiple rounds. Which round of a part is payout relevant will be randomly and with equal probability decided at the end of the experiment (part III). Since you do not know which round will be drawn, it is optimal to behave as if every round is payout-relevant.

At the end of the experiment your points will be converted into Euro and immediately paid out to you in cash. For showing up on time you receive EUR 4 in addition to what you will earn in the experiment.

Anonymity

The analysis of the experiment will be anonymous. That is, we will never link your name with the data generated in the experiment. You will not lear n the identity of any other participant, neither before nor after the experiment. Also the other partic ipants will not learn your identity. At the end of the experiment, you have to sign a receipt to conf irm the payments you received. This receipt will only be used for accounting purposes.

Part I

Task

In this part, we ask you to forecast the the price Y of a fictitious stock. To do this, you receive two values W_1 and W_2, which underlie the price of the stock. You will not learn how exactly the price of the stock is formed out of the two values and a possible constant. However, you will receive examples for this relation, which **will not change** throughout the experiment. Please enter the predicted price of the stock into the respective window on the screen and click on OK. You have 60 seconds for this task. There are no advantages or disadvantages if you enter your solution faster than 60 seconds. You cannot change your input after clicking on OK. You can enter integer values between 1 and 500.

Procedure

At the beginning of the experiment you receive 100 points. 10 points are equal to EUR 1. To get a feeling for the relationship of the stock with the two values, you will once receive 10 examples at the beginning of the experiment on a piece of paper. You then have 5 minutes to study these examples. You can keep them for the rest of the experiment, but may not leave with them.

Next, you have the possibility to practice the task. There are 5 practice rounds with 60 seconds time each. After the five practice rounds you will be shown the true price of the stock, your forecast and the deviation of your forecast. The practice rounds do not influence your payout, but should help you in estimating your abilities for this task.

After the practice rounds the task will be done ten more times. This time, the accuracy of your forecast influences your payout. Every unit that your forecast deviates from the true value leads to a reduction of 1 point.

At the end of the experiment, one out of the 10 rounds will be chosen randomly and with equal probability. The forecasting error from this chosen round will be deducted from your 100 points. If the error is larger or equal to 100 points, you receive no payout from this part.

Insurance

Before solving the task, you have the possibility to buy an insurance. This insurance costs you once 22.5 points and is valid for all 10 rounds. The insurance reimburses 65% of your forecasting error. This means that, if you own the insurance, only 35% of your forecasting error will be deducted from your points.

However, it is not sure if you receive the insurance. In a first step you have to indicate if you want to buy the insurance. If you want to buy the insurance, you will actually receive it with a probability of 70%. With a probability of 30% you will not receive it. In this case you also don't need to pay 22.5 points. The reverse holds, if you indicate that you do not want to buy the insurance. With a probability of 70% you will not receive it, and with a probability of 30% you will receive it nevertheless and you have to pay 22.5 points.

After you decided for or against the purchase of the insurance, you will be informed if you received it or not. Then the 10 rounds start. Only at the end of the experiment will you know the correct value, your forecast and the deviation of your forecast. None of the other participants will ever be informed about your forecast, your choice or receipt of the insurance.

When choosing the insurance, you can activate a calculator by clicking on it symbol in the lower right corner on the screen.

Payment

The payout-relevant round will be drawn at the end of the experiment. If you did not receive an insurance, profit from this part of the experiment will be

$$(100 - |PriceStock - Forecast|) \times 0.1\text{EUR}.$$

If you did receive the insurance your profit will be

$$(100 - |PriceStock - Forecast| \times 35\% - 22.5) \times 0.1\text{EUR}.$$

If you do not enter any forecast within 60 seconds in a round and if this round is chosen as payout-relevant you do not receive any profit from this part of the experiment, even if you have the insurance.

Let's look at some examples.

Example 1

After the practice rounds you decide against buying the insurance. You receive the message that you actually did not get the insurance. Now you perform the task 10 times. At the end of the experiment a random draw decides that round 7 is payout relevant. The true price of the stock in this round was 122. Your prediction was 170. The absolute difference of 48 will be deducted from your 100 points. Converted to euros you will receive $(100 - 48) \times 0.) = 5.2$ Euro.

Example 2

After the practice rounds you decide to buy the insurance. You receive the message that you actually did get the insurance. Now you perform the task 10 times. At the end of the experiment a random draw decides that round 2 is payout relevant. The true price of the stock in this round was 99. Your prediction was 105, so your forecasting error equals 6. The insurance reimburses 65% of your error, or 3.9 points which will be rounded to 4. Hence, only 2 points will be deducted from your 100 points. However the price of the insurance of 22.5 points will also be deducted. Converted to euros you will receive $(100 - 6 \times 35\% - 22.5) \times 0.1 = 7.6$ Euro.

Example 3

After the practice rounds you decide to buy the insurance. However you receive the message that you did not get the insurance. Now you perform the task 10 times.

At the end of the experiment a random draw decides that round 10 is payout relevant. The true price of the stock in this round was 150. Your prediction was 100. Since you did not get the insurance a full 50 points will be deducted from your 100 points. Converted to euros you will receive $(100 - 50) \times 0.1 = 5$ Euro.

Example 4

After the practice rounds you decide against buying the insurance. However you receive the message that you did get the insurance. Now you perform the task 10 times. At the end of the experiment a random draw decides that round 3 is payout relevant. The true price of the stock in this round was 175. Your prediction was 125, so your forecasting error equals 50. The insurance reimburses 65% of your error, or 32.5 points which will be rounded to 33. Hence, only 17 points will be deducted from your initial 100 points. However the price of the insurance of 22.5 points will also be deducted. Converted to euros you will receive $(100 - 50 \times 35\% - 22.5) \times 0.1 = 6.1$ Euro.

Examples for Part I

Here you find 10 examples on the relation of the fictitious stock Y and the two values W_1 and W_2. The exact form of this relationship is identical in the examples, the practice rounds and the payoff-relevant rounds.

Y	W_1	W_2
137	73	95
160	152	85
175	79	152
151	100	87
115	76	49
85	27	37
212	219	139
129	244	7
203	14	217
90	69	25

Please leave this paper on the table when you exit the room.

References

Agüero, J. M. and M. S. Marks (2008). Motherhood and female labor force participation: evidence from infertility shocks. *American Economic Review 98*(2), 500–504.

Akerlof, G. A. (1997). Social distance and social decisions. *Econometrica 65*(5), 1005–1027.

Altonji, J. G., T. E. Elder, and C. R. Taber (2005). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of Political Economy 113*(1), 151.

Ambrasat, J. and J. Schupp (2011). Handgreifkraftmessung im sozio-oekonomischen panel (soep) 2006 und 2008. Technical report, DIW Berlin, German Institute for Economic Research.

Angrist, J., V. Lavy, and A. Schlosser (2010). Multiple experiments for the causal link between the quantity and quality of children. *Journal of Labor Economics 28*(4), 773–824.

Angrist, J. D. (1990). Lifetime earnings and the vietnam era draft lottery: evidence from social security administrative records. *American Economic Review*, 313–336.

Angrist, J. D. and W. N. Evans (1998). Children and their parents' labor supply: Evidence from exogenous variation in family size. *American Economic Review 88*(3), 450 – 477.

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association 91*(434), 444–455.

Angrist, J. D. and J.-S. Pischke (2008). *Mostly harmless econometrics: An empiricist's companion.* Princeton university press.

Angrist, J. D. and J.-S. Pischke (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives 24*(2), 3–30.

Ashley, R. (2009). Assessing the credibility of instrumental variables inference with imperfect instruments via sensitivity analysis. *Journal of Applied Econometrics 24*(2), 325–337.

Åslund, O. and H. Grönqvist (2010). Family size and child outcomes: Is there really no trade-off? *Labour Economics 17*(1), 130–139.

Baker, D., H. Köhler, and M. Stock (2007). Socialist ideology and the contraction of higher education: Institutional consequences of state manpower and education planning in the former east germany. *Comparative Education Review 51*(3), 353–377.

Baumeister, R. F. (1984). Choking under pressure: self-consciousness and paradoxical effects of incentives on skillful performance. *Journal of Personality and Social Psychology 46*(3), 610.

Becker, G. S. and H. G. Lewis (1973). On the interaction between the quantity and quality of children. *Journal of Political Economy 81*(2), 279–88.

Becker, G. S. and N. Tomes (1994). Human capital and the rise and fall of families. In *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education (3rd Edition)*, pp. 257–298. The University of Chicago Press.

Benoît, J.-P. and J. Dubra (2011). Apparent overconfidence. *Econometrica 79*(5), 1591–1625.

Bergh, C. (2005). Single embryo transfer: a mini-review. *Human Reproduction 20*(2), 323–327.

Berkman, L. F., Y. Zheng, M. M. Glymour, M. Avendano, A. Börsch-Supan, and E. L. Sabbath (2015). Mothering alone: cross-national comparisons of later-life disability and health among women who were single mothers. *Journal of Epidemiological Community Health 69*(9), 865–872.

Bisbee, J., R. Deheijia, C. Pop-Eleches, and C. Samii (2017). Local instruments, global extrapolation: External validity of the same-sex-fertility-labor supply local average treatment effect. *Journal of Labor Economics*, forthcoming.

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*(2), 669 – 700.

Black, S. E., P. J. Devereux, and K. G. Salvanes (2007). From the cradle to the labor market? the effect of birth weight on adult outcomes. *Quarterly Journal of Economics 122*(1), 409–439.

Black, S. E., P. J. Devereux, and K. G. Salvanes (2010). Small family, smart family? family size and the iq scores of young men. *Journal of Human Resources 45*(1), 33–58.

Bloom, D. E., D. Canning, and J. Sevilla (2004). The effect of health on economic growth: a production function approach. *World Development 32*(1), 1–13.

Bó, P. D., A. Foster, and L. Putterman (2010). Institutions and behavior: Experimental evidence on the fffects of democracy. *American Economic Review 100*(5), 2205–2229.

Bomsel-Helmreich, O. and W. Al Mufti (2005). The phenomenon of monozygosity: Spontaneous zygotic splitting. *Multiple pregnancy: Epidemiology, Gestation & perinatal outcome*, 94–101.

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

Börsch-Supan, A. (2003). Labor market effects of population aging. *Labour 17*(1), 5–44.

Bortolus, R., F. Parazzini, L. Chatenoud, G. Benzi, M. M. Bianchi, and A. Marini (1999). The epidemiology of multiple births. *Human Reproduction Update 5*(2), 179–187.

Braakmann, N. and J. Wildman (2016). Reconsidering the effect of family size on labour supply: the twin problems of the twin birth instrument. *Journal of the Royal Statistical Society, Series A 179*(4), 1093–1115.

Bronars, S. G. and J. Grogger (1994). The economic consequences of unwed motherhood: Using twin births as a natural experiment. *American Economic Review 84*(5), 1141.

Brown, P. M. (1998). Experimental evidence on the importance of competing for profits on forecasting accuracy. *Journal of Economic Behavior & Organization 33*(2), 259–269.

Browne, M. J., C. Knoller, and A. Richter (2015). Behavioral bias and the demand for bicycle and flood insurance. *Journal of Risk and Uncertainty 50*(2), 141–160.

Buhling, K. J., W. Henrich, E. Starr, M. Lubke, S. Bertram, G. Siebert, and J. W. Dudenhausen (2003). Risk for gestational diabetes and hypertension for women with twin pregnancy compared to singleton pregnancy. *Archives of gynecology and obstetrics 269*(1), 33–36.

Bulmer, M. G. (1976). Is weinberg's method valid? *Acta geneticae medicae et gemellologiae 25*, 25–28.

Butcher, K. F. and A. Case (1994). The effect of sibling sex composition on women's education and earnings. *Quarterly Journal of Economics 109*(3), 531–563.

Bütikofer, A. (2010). Sibling sex composition and cost of children. Mimeo.

Cáceres-Delpiano, J. (2006). The impacts of family size on investment in child quality. *Journal of Human Resources 41*(4), 738–754.

Cáceres-Delpiano, J. and M. Simonsen (2012). The toll of fertility on mothers' wellbeing. *Journal of Health Economics 31*(5), 752–766.

Caliendo, M., D. A. Cobb-Clark, and A. Uhlendorff (2015). Locus of control and job search strategies. *Review of Economics and Statistics 97*(1), 88–103.

Camerer, C. and D. Lovallo (1999). Overconfidence and excess entry: An experimental approach. *American Economic Review 89*(1), 306–318.

Cattell, R. B. (1987). *Intelligence: Its Structure, Growth and Action.* Elsevier.

Chan, J. C. C. and J. L. Tobias (2015). Priors and posterior computation in linear endogenous variable models with imperfect instruments. *Journal of Applied Econometrics 30*(4), 650–674.

Clark, J. and L. Friesen (2009). Overconfidence in forecasts of own performance: An experimental study. *Economic Journal 119*(534), 229–251.

Conley, T. G., C. B. Hansen, and P. E. Rossi (2012). Plausibly exogenous. *Review of Economics and Statistics 94*(1), 260–272.

Coren, S. (2012). *The left-hander syndrome: The causes and consequences of left-handedness*. Simon and Schuster.

Coren, S. and D. F. Halpern (1991). Left-handedness: A marker for decreased survival fitness. *Psychological Bulletin 109*(1), 90.

Coudé, F., C. Mignot, S. Lyonnet, and A. Munnich (2006). Discontinuity in the fall of left-handedness in a french population: A may'68 effect? *Laterality 11*(1), 33–35.

Cunha, F. and J. Heckman (2007). The technology of skill formation. *American Economic Review 97*(2), 31–47.

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

Cunha, F. and J. J. Heckman (2010). Investing in our young people. nber working paper no. 16201. *National Bureau of Economic Research*.

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

Currie, J. and D. Almond (2011). Human capital development before age five. *Handbook of Labor Economics 4*, 1315–1486.

De la Rosa, L. E. (2011). Overconfidence and moral hazard. *Games and Economic Behavior 73*(2), 429–451.

Dehne, M. and J. Schupp (2007). Persönlichkeitsmerkmale im sozio-oekonomischen panel (soep)-konzept, umsetzung und empirische eigenschaften. *Research Notes 26*, 1–70.

Del Boca, D. (2002). The effect of child care and part time opportunities on participation and fertility decisions in italy. *Journal of Population Economics 15*(3), 549–573.

Denny, K. and V. O'Sullivan (2007). The economic consequences of being left-handed: Some sinister results. *Journal of Human Resources 42*(2), 353–374.

Dohmen, T. and A. Falk (2011). Performance pay and multidimensional sorting: Productivity, preferences, and gender. *American Economic Review 101*(2), 556–590.

Dohmen, T., A. Falk, D. Huffman, and U. Sunde (2010). Are risk aversion and impatience related to cognitive ability? *American Economic Review 100*(3), 1238–60.

Ekbom, A. (2011). The swedish multi-generation register. *Methods in Biobanking 675*, 215–220.

Evans, W. N. and T. J. Moore (2012). Liquidity, economic activity, and mortality. *Review of Economics and Statistics 94*(2), 400–418.

Fauser, B. C., P. Devroey, and N. S. Macklon (2005). Multiple birth resulting from ovarian stimulation for subfertility treatment. *The Lancet 365*(9473), 1807–1816.

Fellman, J. and A. W. Eriksson (2006). Weinberg's differential rule reconsidered. *Human Biology 78*(3), 253–275.

Figlio, D., J. Guryan, K. Karbownik, and J. Roth (2014). The effects of poor neonatal health on children's cognitive development. *American Economic Review 104*(12), 3921–3955.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental economics 10*(2), 171–178.

Frijters, P., D. W. Johnston, M. Shah, and M. A. Shields (2009). To work or not to work? child development and maternal labor supply. *American Economic Journal: Applied Economics 1*(3), 97–110.

Frijters, P., D. W. Johnston, M. Shah, and M. A. Shields (2013). Intrahousehold resource allocation: Do parents reduce or reinforce child ability gaps? *Demography 50*(6), 2187–2208.

Gächter, S. and E. Renner (2010). The effects of (incentivized) belief elicitation in public goods experiments. *Experimental Economics 13*(3), 364–377.

Ganderton, P. T., D. S. Brookshire, M. McKee, S. Stewart, and H. Thurston (2000). Buying insurance for disaster-type risks: experimental evidence. *Journal of Risk and Uncertainty 20*(3), 271–289.

Goldin, C. and J. Mitchell (2017). The new life cycle of women's employment: Disappearing humps, sagging middles, expanding tops. *Journal of Economic Perspectives 31*(1), 161–182.

Goodman, J. (2014). The wages of sinistrality: Handedness, brain structure, and human capital accumulation. *Journal of Economic Perspectives 28*(4), 193–212.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with orsee. *Journal of the Economic Science Association 1*(1), 114–125.

Gronau, R. (1973). The effect of children on the housewife's value of time. *Journal of Political Economy 81*(2), 168–199.

Grossman, M. (1972). On the concept of health capital and the demand for health. *Journal of Political economy 80*(2), 223–255.

Groves, M. O. (2005). How important is your personality? labor market returns to personality for women in the us and uk. *Journal of Economic Psychology 26*(6), 827–841.

Grundy, E. and Ø. Kravdal (2010). Fertility history and cause-specific mortality: a register-based analysis of complete cohorts of norwegian women and men. *Social Science & Medicine 70*(11), 1847–1857.

Hall, J. G. (2003). Twinning. *The Lancet 362*(9385), 735–743.

Halpern, D. and S. Coren (1988). Do right-handers live longer? *Nature 333*(6170), 213–213.

Hank, K. and B. Schaan (2008). Cross-national variations in the correlation between frequency of prayer and health among older europeans. *Research on Aging 30*(1), 36–54.

Hanushek, E. A. and L. Woessmann (2008). The role of cognitive skills in economic development. *Journal of Economic Literature 46*(3), 607–668.

Hardin, J., S. Selvin, S. L. Carmichael, and G. M. Shaw (2009). The estimated probability of dizygotic twins: A comparison of two methods. *Twin Research and Human Genetics 12*(1), 79–85.

Harris, L. J. (1980). Left-handedness: Early theories, facts, and fancies. *Neuropsychology of Left-handedness*, 3–78.

Harris, L. J. (1990). Cultural influences on handedness: Historical and contemporary theory and evidence. *Advances in Psychology 67*, 195–258.

Harris, L. J. (1993). Do left-handers die sooner than right-handers? commentary on coren and halpern's (1991) left-handedness: a marker for decreased survival fitness. *Psychological bulletin 114*(2), 203–234.

Heckman, J. (2000). Policies to foster human capital. *Research in economics 54*(1), 3–56.

Heckman, J., R. Pinto, and P. Savelyev (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review 103*(6), 2052–2086.

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

Heckman, J. J. and J. R. Walker (1990). The relationship between wages and income and the timing and spacing of births: evidence from swedish longitudinal data. *Econometrica*, 1411–1441.

Heineck, G. and S. Anger (2010). The returns to cognitive abilities and personality traits in germany. *Labour Economics 17*(3), 535–546.

Hoelzl, E. and A. Rustichini (2005). Overconfident: Do you put your money on it? *The Economic Journal 115*(503), 305–318.

Hurt, L. S., C. Ronsmans, and S. L. Thomas (2006). The effect of number of births on women's mortality: systematic review of the evidence for women who have completed their childbearing. *Population studies 60*(1), 55–71.

Hurtado, D. A., C. A. Okechukwu, O. M. Buxton, L. Hammer, G. C. Hanson, P. Moen, L. C. Klein, and L. F. Berkman (2016). Effects on cigarette consumption of a work–family supportive organisational intervention: 6-month results from the work, family and health network study. *Journal of Epidemiology and Community Health 70*(12), 1155–1161.

Imai, K., L. Keele, and T. Yamamoto (2010). Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science 25*(1), 51–71.

Jacobsen, J. P., J. W. Pearce III, and J. L. Rosenbloom (1999). The effects of childbearing on married women's labor supply and earnings. *Journal of Human Resources 34*(3), 449 – 474.

Jaumotte, F. (2004). Labour force participation of women. *OECD Economic studies 2003*(2), 51–108.

Johnston, D. W., M. E. Nicholls, M. Shah, and M. A. Shields (2009). Nature's experiment? handedness and early childhood development. *Demography 46*(2), 281–301.

Johnston, D. W., M. E. Nicholls, M. Shah, and M. A. Shields (2013). Handedness, health and cognitive development: Evidence from children in the national longitudinal survey of youth. *Journal of the Royal Statistical Society: Series A 176*(4), 841–860.

Karlan, D. and J. Zinman (2009). Observing unobservables: Identifying information asymmetries with a consumer credit field experiment. *Econometrica 77*(6), 1993–2008.

Kesternich, I., B. Siflinger, J. P. Smith, and J. K. Winter (2014). The effects of world war ii on economic and health outcomes across europe. *Review of Economics and Statistics 96*(1), 103–118.

Kiecolt-Glaser, J. K., K. J. Preacher, R. C. MacCallum, C. Atkinson, W. B. Malarkey, and R. Glaser (2003). Chronic stress and age-related increases in the proinflammatory cytokine il-6. *Proceedings of the National Academy of Sciences 100*(15), 9090–9095.

Klöppel, S., J.-F. Mangin, A. Vongerichten, R. S. Frackowiak, and H. R. Siebner (2010). Nurture versus nature: Long-term impact of forced right-handedness on structure of pericentral cortex and basal ganglia. *The Journal of Neuroscience 30*(9), 3271–3275.

Kőszegi, B. (2014). Behavioral contract theory. *Journal of Economic Literature 52*(4), 1075–1118.

Kraay, A. (2012). Instrumental variables regressions with uncertain exclusion restrictions: A bayesian approach. *Journal of Applied Econometrics 27*(1), 108–128.

Krueger, A. B. and J.-S. Pischke (1995). A comparative analysis of east and west german labor markets: Before and after unification. In *Differences and Changes in Wage Structures*, pp. 405–446. University of Chicago Press.

Kruk, K. E. and S. Reinhold (2014). The effect of children on depression in old age. *Social Science & Medicine 100*, 1–11.

Kulkarni, A. D., D. J. Jamieson, H. W. Jones Jr, D. M. Kissin, M. F. Gallo, M. Macaluso, and E. Y. Adashi (2013). Fertility treatments and multiple births in the united states. *New England Journal of Medicine 369*(23), 2218–2225.

Kushner, H. I. (2011). Retraining the king's left hand. *The Lancet 377*(9782), 1998–1999.

Kushner, H. I. (2012). Retraining left-handers and the aetiology of stuttering: the rise and fall of an intriguing theory. *Laterality: Asymmetries of Body, Brain and Cognition 17*(6), 673–693.

Kushner, H. I. (2013). Why are there (almost) no left-handers in china? *Endeavour 37*(2), 71–81.

Lalive, R. and J. Zweimüller (2009). How does parental leave affect fertility and return to work? evidence from two natural experiments. *The Quarterly Journal of Economics 124*(3), 1363–1402.

Lang, F. R., D. Weiss, A. Stocker, and B. von Rosenbladt (2007). Assessing cognitive capacities in computer-assisted survey research: Two ultra-short tests of intellectual ability in the german socio-economic panel (soep). *Schmollers Jahrbuch 127*(183), 192.

Larkin, I. and S. Leider (2012). Incentive schemes, sorting, and behavioral biases of employees: Experimental evidence. *American Economic Journal: Microeconomics 4*(2), 184–214.

Laury, S. K., M. M. McInnes, and J. T. Swarthout (2009). Insurance decisions for low-probability losses. *Journal of Risk and Uncertainty 39*(1), 17–44.

Lee, J. (2008). Sibling size and investment in children's education: An asian instrument. *Journal of Population Economics 21*(4), 855–875.

Lindbeck, A. and S. Nyberg (2006). Raising children to work hard: Altruism, work norms, and social insurance. *Quarterly Journal of Economics 121*(4), 1473–1503.

Low, C. A., R. C. Thurston, and K. A. Matthews (2010). Psychosocial factors in the development of heart disease in women: current research and future directions. *Psychosomatic medicine 72*(9), 842.

Lundborg, P., E. Plug, and A. W. Rasmussen (2017). Can women have children and a career? iv evidence from ivf treatments. *American Economic Review*, forthcoming.

MacGillivray, I., M. Samphier, and J. Little (1988). Factors affecting twinning. In I. MacGillivray, D. Campbell, and B. Thompson (Eds.), *Twinning and twins*. John Wiley & Sons.

Malmendier, U. and G. Tate (2005). Ceo overconfidence and corporate investment. *The Journal of Finance 60*(6), 2661–2700.

Martin, J. A., B. E. Hamilton, M. J. Osterman, N. C. for Health Statistics (US), et al. (2012). Three decades of twin births in the united states, 1980-2009.

McCrae, R. R. and P. T. Costa Jr (1999). A five-factor theory of personality. In *Handbook of Personality: Theory and Research*, pp. 139. Elsevier.

McManus, I., A. Davison, and J. A. Armour (2013). Multilocus genetic models of handedness closely resemble single-locus models in explaining family data and are compatible with genome-wide association studies. *Annals of the New York Academy of Sciences 1288*(1), 48–58.

McManus, I. C. (2009). The history and geography of human handedness. In *Language lateralization and psychosis*, pp. 37–57. Cambridge University Press Cambridge.

Medland, S., I. Perelle, V. De Monte, and L. Ehrman (2004). Effects of culture, sex, and age on the distribution of handedness: an evaluation of the sensitivity of three measures of handedness. *Laterality 9*(3), 287–297.

Merkle, C. and M. Weber (2011). True overconfidence: The inability of rational information processing to account for apparent overconfidence. *Organizational Behavior and Human Decision Processes 116*(2), 262–271.

Mogstad, M. and M. Wiswall (2016). Testing the quantity-quality model of fertility: Estimation using unrestricted family size models. *Quantitative Economics 7*(1), 157–192.

Moore, D. A. and D. M. Cain (2007). Overconfidence and underconfidence: When and why people underestimate (and overestimate) the competition. *Organizational Behavior and Human Decision Processes 103*(2), 197–213.

Moore, D. A. and P. J. Healy (2008). The trouble with overconfidence. *Psychological Review 115*(2), 502–517.

Murad, Z., M. Sefton, and C. Starmer (2016). How do risk attitudes affect measured confidence? *Journal of Risk and Uncertainty 52*(1), 21–46.

Nelson, C. C., Y. Li, G. Sorensen, and L. F. Berkman (2012). Assessing the relationship between work–family conflict and smoking. *American Journal of Public Health 102*(9), 1767–1772.

Nevo, A. and A. M. Rosen (2012). Identification with imperfect instruments. *Review of Economics and Statistics 94*(3), 659–671.

Ohse, M.-D. (2010). *Jugend nach dem Mauerbau: Anpassung, Protest und Eigensinn (DDR 1961–1974).* Ch. Links Verlag.

Papadatou-Pastou, M., M. Martin, M. R. Munafo, and G. V. Jones (2008). Sex differences in left-handedness: A meta-analysis of 144 studies. *Psychological Bulletin 134*(5), 677.

Perelle, I. B. and L. Ehrman (1994). An international study of human handedness: The data. *Behavior Genetics 24*(3), 217–227.

Perelle, I. B. and L. Ehrman (2005). On the other hand. *Behavior genetics 35*(3), 343–350.

Piatek, R. and P. Pinger (2015). Maintaining (locus of) control? data combination for the identification and inference of factor structure models. *Journal of Applied Econometrics 31*(4), 734–755.

Pison, G. and A. V. D'Addato (2006). Frequency of twin births in developed countries. *Twin Research and Human Genetics 9*(2), 250–259.

Porac, C., S. Coren, and A. Searleman (1986). Environmental factors in hand preference formation: Evidence from attempts to switch the preferred hand. *Behavior Genetics 16*(2), 251–261.

Porac, C. and W. L. B. Martin (2007). A cross-cultural comparison of pressures to switch left-hand writing: Brazil versus canada. *Laterality 12*(3), 273–291.

Porac, C. and A. Searleman (2002). The effects of hand preference side and hand preference switch history on measures of psychological and physical well-being and cognitive performance in a sample of older adult right- and left-handers. *Neuropsychologia 40*(12), 2074–2083.

Rauh-Hain, J. A., S. Rana, H. Tamez, A. Wang, B. Cohen, A. Cohen, F. Brown, J. L. Ecker, S. A. Karumanchi, and R. Thadhani (2009). Risk for developing gestational diabetes in women with twin pregnancies. *Journal of Maternal-Fetal and Neonatal Medicine 22*(4), 293–299.

Reddy, U. M., A. M. Branum, and M. A. Klebanoff (2005). Relationship of maternal body mass index and height to twinning. *American College of Obstetricians and Gynecologists 105*(3), 593–597.

Ridker, P. M., C. H. Hennekens, J. E. Buring, and N. Rifai (2000). C-reactive protein and other markers of inflammation in the prediction of cardiovascular disease in women. *New England Journal of Medicine 342*(12), 836–843.

Robles, T. F., R. Glaser, and J. K. Kiecolt-Glaser (2005). Out of balance: A new look at chronic stress, depression, and immunity. *Current Directions in Psychological Science 14*(2), 111–115.

Rosenzweig, M. R. and K. I. Wolpin (1980a). Life-cycle labor supply and fertility: Causal inferences from household models. *Journal of Political Economy 88*(2), 328–348.

Rosenzweig, M. R. and K. I. Wolpin (1980b). Testing the quantity-quality fertility model: The use of twins as a natural experiment. *Econometrica 48*(1), 227–240.

Rosenzweig, M. R. and K. I. Wolpin (2000). Natural" natural experiments" in economics. *Journal of Economic Literature 38*(4), 827–874.

Rothschild, M. and J. Stiglitz (1976). Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *Quarterly Journal of Economics 90*(4), 629–649.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General and Applied 80*(1), 1–28.

Ruebeck, C. S., J. E. Harrington Jr, and R. Moffitt (2007). Handedness and earnings. *Laterality 12*(2), 101–120.

Sabbath, E. L., I. M. Guevara, M. M. Glymour, and L. F. Berkman (2015). Use of life course work–family profiles to predict mortality risk among us women. *American Journal of Public Health 105*(4), 96–102.

Sabbath, E. L., I. Mejía-Guevara, C. Noelke, and L. F. Berkman (2015). The long-term mortality impact of combined job strain and family circumstances: A life course analysis of working american mothers. *Social Science & Medicine 146*, 111–119.

Sandroni, A. and F. Squintani (2007). Overconfidence, insurance, and paternalism. *American Economic Review 97*(5), 1994–2004.

Sandroni, A. and F. Squintani (2013). Overconfidence and asymmetric information: The case of insurance. *Journal of Economic Behavior & Organization 93*, 149–165.

Santos-Pinto, L. (2008). Positive self-image and incentives in organisations. *The Economic Journal 118*(531), 1315–1332.

Sattler, J. B. (1996). *Der umgeschulte Linkshänder oder der Knoten im Gehirn.* Auer.

Satz, P. (1972). Pathological left-handedness: An explanaory model. *Cortex 8*(2), 121–135.

Sautmann, A. (2011). Contracts for agents with biased beliefs: Some theory and an experiment. *American Economic Journal: Microeconomics*.

Scheinkman, J. A. and W. Xiong (2003). Overconfidence and speculative bubbles. *Journal of Political Economy 111*(6), 1183–1220.

Schönberg, U. and J. Ludsteck (2014). Expansions in maternity leave coverage and mothers? labor market outcomes after childbirth. *Journal of Labor Economics 32*(3), 469–505.

Schram, A. and J. Sonnemans (2011). How individuals choose health insurance: An experimental analysis. *European Economic Review 55*(6), 799–819.

Shavell, S. (1979). On moral hazard and insurance. *Quarterly Journal of Economics*, 541–562.

Smith, J. P. (1999). Healthy bodies and thick wallets: the dual relation between health and economic status. *Journal of Economic Perspectives 13*(2), 144–166.

So, T., P. Brown, A. Chaudhuri, D. Ryvkin, and L. Cameron (2017). Piece-rates and tournaments: Implications for learning in a cognitively challenging task. *Journal of Economic Behavior & Organization 142*, 11–23.

Staiger, D. and J. H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica 65*(3), 557–586.

Sutter, M., S. Haigner, and M. G. Kocher (2010). Choosing the carrot or the stick? endogenous institutional choice in social dilemma situations. *The Review of Economic Studies 77*(4), 1540–1566.

Sweden, S. (2010). Household budget survey (hbs) 2007-2009, expenditures and income.

Thaler, R. H. (2000). From homo economicus to homo sapiens. *Journal of Economic Perspectives 14*(1), 133–141.

Thurin, A., J. Hausken, T. Hillensjö, B. Jablonowska, A. Pinborg, A. Strandell, and C. Bergh (2004). Elective single-embryo transfer versus double-embryo transfer in in vitro fertilization. *New England Journal of Medicine 351*(23), 2392–2402.

Tong, S. and R. Short (1998). Dizygotic twinning as a measure of human fertility. *Human Reproduction 13*(1), 95–98.

Trautmann, S. T. and G. Kuilen (2015). Belief elicitation: A horse race among truth serums. *The Economic Journal 125*(589), 2116–2135.

Troske, K. R. and A. Voicu (2012). The effect of the timing and spacing of births on the level of labor market involvement of married women. *Empirical Economics*, 1–39.

Van Hedel, K., I. Mejía-Guevara, M. Avendaño, E. L. Sabbath, L. F. Berkman, J. P. Mackenbach, and F. J. van Lenthe (2016). Work–family trajectories and the higher cardiovascular risk of american women relative to women in 13 european countries. *American Journal of Public Health 106*(8), 1449–1456.

Vandegrift, D. and P. Brown (2003). Task difficulty, incentive effects, and the selection of high-variance strategies: an experimental examination of tournament behavior. *Labour Economics 10*(4), 481–497.

Vlietinck, R., C. Derom, and R. Derom (1988). The validity of weinberg's rule in the east flanders prospective twin survey (efpts). *Acta geneticae medicae et gemellologiae 37*, 137–141.

Wagner, G. G., J. R. Frick, and J. Schupp (2007). The german socio-economic panel study (soep)–scope, evolution and enhancements. *Schmollers Jahrbuch 127*, 139–169.

Weinberg, W. (1901). Beiträge zur physiologie und pathologie der mehrlingsgeburten beim menschen. *Pflügers Archiv - European Journal of Physiology 88*(6-8), 346–430.

Witelson, S. F. (1985). The brain connection: The corpus callosum is larger in left-handers. *Science 229*(4714), 665–668.

Young, H. P. (1996). The economics of convention. *Journal of Economic Perspectives 10*(2), 105–122.

Zverev, Y. P. (2006). Cultural and environmental pressure against left-hand preference in urban and semi-urban malawi. *Brain and Cognition 60*(3), 295–303.