Cautious Inference: Random Life Course Events of Parents and Children in Context

DOCTORAT EN CIÈNCIES POLÍTIQUES I SOCIALS

TESI DOCTORAL UPF 2014

Sander Wagner

Directors de la tesi:

Dr. Gøsta ESPING-ANDERSEN[1]

[1]Departament de Ciències Polítiques i Socials
Universitat Pompeu Fabra
Spain

Thesis committee:

Prof. Fabrizio Bernardi
Department of Political and Social Sciences, European University Institute, Italy

Prof. Marianne Simonsen
Department of Economics and Business, Aarhus University, Denmark

Prof. Maria José González
Department of Political and Social Sciences, Universitat Pompeu Fabra, Spain

Dedication

To Laura Garcia Matos and to my parents.

Acknowledgements

First and foremost I want to thank my thesis advisor Gosta Esping-Andersen for his firm support, his brilliant intellectual guidance and above all for his patience throughout the years in which I worked on this thesis. Thanks for discussing ideas and helping me on various aspects of the thesis have to go out to practically the entire department of social and political sciences at Universitat Pompeu Fabra. At the risk of forgetting someone important I want to extend special thanks for helping me out with various aspects of this thesis and for important ideas or discussions to Diederik Boertien, Lea Pessin, Bruno Arpino, Goncalo Alves-Pina, Mette Gortz, Marianne Simonsen, Stephan Litschig and Angela Fiedler.

Contents

List of Figures

List of Tables

1 Introduction

This thesis has three articles that are tied together by two themes. The first unifying theme is about the subject matters, which the three research papers that are presented here explore. They all analyze certain key events playing out in the life of Danish parents and children and look at how those events influenced the lives of affected individuals in later years.
The second theme is methodological. All three papers have a research design, which aims to draw causal inferences by exploiting random variation in the assignment of key life-course events. It may seem odd that the idea of "randomness", which features in this thesis' title and is commonly defined as "a lack of purpose or predictability" has made its way to the heart of our social science tool case. But when trying to uncover the causal relationships that govern the social world, "purpose" often turns out to be the enemy. Imagine we were interested in finding out the effects of a new preschool initiative on children's later test scores. Now go on to imagine two extremely purposeful mothers, one carefully planning out her child's path through state institutions including preschool and another one setting out a careful plan for teaching it as much as possible in the nest of the home, potentially even trying to homeschool the child. Everything about these two mothers, the family they have, the community they are surrounded by, how they raise their kids etc. is bound to be different. Thus when we find that their kids have different life outcomes, it would be most uncareful of us to attribute those differences to whether the children attended preschool or not. Too many other important factors influencing their life are bound to differ as well. What we instead want is mothers, and more importantly children, that we expect to be similar, with some of them due to a stroke of chance attending pre-school and others not. Sometimes nature, luck or a certain policymakers' design allows us to observe cases, where similar individuals through an event outside of their control end up in different situations and we can thus go on to identify the effects of those situations on them. This is what we will call treatment effects. It is this type of randomness the title alludes to. This thesis is about the great opportunities in terms of statistical identification that random variation

offers us but it is also about how carefully it has to be approached, since few things in this world turn out to be truly random.
In one of the here presented studies we exploit almost random variation in a policy assignment and in the other two we look at what is called "natural experiments", cases in which an event of nature randomly assigns certain conditions to some individuals but not to others.
The three papers of this thesis thus all employ the causal identification techniques that Angrist and Pischke have described as the "credibility revolution in empirical economics" (Angrist and Pischke 2010). Building on a fundamental critique of the econometrics of the 1970s and 80s brought forward by Leamer (Leamer 1983) they argue that, while the big structural models employed at that time failed to ensure proper identification, experimental and quasi-experimental design has significantly restored the trustworthiness of our empirical estimates. While this view is generally embraced there have also been voices of criticism, warning of a too naive embrace and a too blind application of those experimental or quasi-experimental techniques. As Raj Chetty put it "People think about the question less than the method ... so you get weird papers, like sanitation facilities in Native American reservations". Even Leamer himself does not fully embrace the enthusiastic answer to his critique that Angrist and Pischke formulate with their promotion of experimental design. Instead he stresses that we should constantly be aware of the "limits of randomization" and that we should conduct careful sensitivity analysis (Leamer 2010). In particular with respect to the use of instrumental variables he points out that we have to be acutely aware of the limitations to satisfying the underlying assumptions and to look carefully at what the consequences of imperfectly fulfilling the assumptions are. In his words there was a cohort of econometricians who "acted as if it were enough to chant instrumental variable". The work of this thesis tries to take the cautious approach to experimental study design, which Leamer formulated, seriously. Indeed the contribution of the here presented work is not the uncovering of new "experiments" that can be used for causal inference. All sources of random variation that are used for drawing inferences have been previously exploited by other researchers. What the here presented work instead attempts to do, is to approach a series of well-known study designs with the cautious skepticism Leamer advises us to preserve. More precisely it looks at whether careful thinking about randomization in its context can yield new insights . The point, which each of the papers tries to make in a distinct way, is that, having found a source of random variation that

allows us to identify causal effects does in no way free us from engaging in a deep contextual analysis of the subject at hand. To make this point in an even stronger form, the here presented work is born out of a deeply cautious belief that mechanical application of causal inference techniques can not only lead to an insufficient interpretation of the results, but in some cases even to misleading results. When reading the three papers, you will see that in some cases "surrounding causal inference with context" can lead to substantially different results or interpretations of what was initially thought to be a clearly established result, while in some cases it strengthens initial findings and intuitions, thereby assuring us that they stand up to the scrutiny under which we put the underlying assumptions or to the way in which we recalculate the results. While it is often perceived as more valuable and exciting to overturn established wisdom by approaching a problem from a new perspective, putting established results under increased scrutiny and confirming them adds just as much in our quest for understanding the world.

Most doctoral theses are unified by one big question that motivates them. The here presented work started out under a similar premise. The goal was to come up with a series of new findings on the effects that pre-school has on mothers and children. One of the works that was planned to appear in this thesis originally was an investigation of whether the availability of pre-school would make it easier for mothers to cope with having children. The idea was to look at whether the well-established negative effects of twinning on a mothers career would be dampened if childcare was broadly accessible. While working on this study, I became increasingly absorbed by the question of whether I could actually trust my own estimates and what their weaknesses were. In the end I was so convinced of the studys' limitations, that I felt incapable of finishing and properly defending it and instead considered it to be a scientifically more honest exercise to instead explore, the insights which I had developed on the limitations of my results. Thus a twin study became a work on limitations of twin studies and a thesis which was supposed to be on early childcare and its effects on the life course turned into a thesis looking at causal inference in life course studies. All this is to warn the reader that if you expect a standard thesis exploring one topic in-depth the three papers might feel incoherent at times. If you do however understand that the guiding "North Star" turned out to be a question, that invariably returned to my head on every question I approached, namely "Can I trust my results?", then you might see that the three papers are indeed born out of a unifying impulse. You will also see that they are held together by studying similar phenomena under a common overarching question, namely "How is causal inference achieved and potentially improved in life-course studies?"

The first paper of this thesis, entitled "A non-linear Assessment of Preschool Effects" is concerned with the life course of children. The key life event that is analyzed here is whether they did or did not attend a high-quality preschool at age 3. We then go on to look at how this affected their cognitive and non-cognitive development as measured by test scores, when the children were aged 7. The paper thus aims to evaluate the effect of a universal child care program on children's later outcomes. The few studies which exist on universal childcare programs have produced a seeming contradiction in the literature on early life interventions. While a great many studies show strong positive effects of targeted early education programs on child development and adult outcomes (Heckman 2006) the few existing studies on universal child care programs have failed to find similar effects. A possible explanation for this contradiction has been put forward in recent research by Havnes and Mogstad (Havnes and Mogstad 2010). While their evaluation of the mean effects of early education in Norway yielded few significant results they instead decided to look at their data using non-linear estimators. What they found was that in those cases, quite strong effects, that were missed by the mean estimates could be found on the more disadvantaged segments of the population. If early childhood education does indeed have stronger effects on more at-risk children, that would offer a possible reconciliation between the conflicting results found in targeted studies and in evaluations of universal care.
To see whether these non-linear effects were a general feature of universal childcare systems, we decided to revisit one of the two other studies (apart from the Havnes and Mogstad one) looking at universal early education in a causal setup. Work by Simonsen and Datta Gupta (Datta Gupta and Simonsen 2010) identified differing policies at the Danish community (Kommune) level, that led to plausibly exogenous quasi-random variation in the access to high-quality preschool for children. While their work found little to no effect of preschool access on children's cognitive and non-cognitive outcomes, a non-randomized analysis of the effects of preschool on Danish children had found significant non-linear effects (Esping-Andersen et al. 2012), with lower-performing boys being particularly positively affected by pre-school attendance. Thus, looking at whether in Denmark, much like in Norway, mean estimates missed the

actual story of the effects of pre-school seemed a plausible avenue to explore. Using non-linear instrumental variable estimators to reestimate the effects of high quality preschool on children's cognitive and non-cognitive outcomes we found effects that in general differed surprisingly little from the initial non-linear estimates and turned out to be substantially less pronounced than the non-instrumented non-linear estimates. Nevertheless the data seemed to indicate that indeed lower-performing boys do profit more from access to high-quality preschool. A surprising result was that lower-performing girls seem to be negatively affected by the attendance of pre-school.

The second paper, entitled "Rusty Instruments? Revisiting the Twin Approach to Estimating the Relationship between Fertility and Maternal Labour Market Outcomes" is concerned with parents, more precisely with how children affect paternal labor-market outcomes. It follows the footsteps of many seminal studies on the subject that used twinning as a source of exogenous variation in the number of children (Rosenzweig and Wolpin 1980). These studies use the occurrence of a twin birth as a randomly occurring event, which increases the number of children a mother has. There are two ways in which the regressions exploiting this random variation can be set up. Researchers either include twinning directly as an explanatory variable in simple least squared regressions or they use it as an instrument for the total number of children, which is then included as an instrumented explanatory variable in the regression. In either case the random variation in the number of children, resulting from a twin birth is exploited to estimate effects on maternal and/or paternal labor-market outcomes. An aspect that is generally acknowledged, but not given much attention is that mothers who twinned tend to have a very different fertility behavior after giving birth, than mothers who did not twin. Given that they have a greater number of children already, twinning mothers tend to get less children after having given birth, than their non-twinning counterparts. We argue that this difference in subsequent fertility behavior has the potential to substantially contaminate any estimates derived from the application of twinning as an explanatory variable in OLS or as an instrument. We show that as a consequence of the differences in subsequent fertility behavior, estimates of the coefficient capturing the effect that a child has on his parents are bound to be upwardly biased . This is particularly true for the effects of children on mothers' careers. We then go on to identify cases in which the subsequent fertility behavior of twinning and non-twinning mothers is much more similar than in the cases normally used for estimation. When comparing mothers giving birth to twins or singletons at a high birth parity, for example at third or fourth birth, the subsequent fertility behavior tends to differ far less than at lower parities. Also, when having a first child or first twin respectively at a high age we find a less pronounced difference in subsequent fertility behavior, which is logical since there is less time available to go on having children. In all of these cases our estimates for the effects of children on mothers show negative labor market effects that are several orders of magnitude bigger than both our initial estimates using traditional estimation techniques and than the findings that have been established in the previous literature. Finally we go on to check the robustness of our estimates in several important ways. To ensure that the rise in in-vitro fertilization, which leads to higher twinning probabilities, does not affect our estimates we reconduct the entire analysis with a sample that does not include births after 1986. This means that the technology was not yet widely available. We find that our results do not differ. We also construct a second counterfactual case in which the subsequent fertility difference between twinning and non-twinning mothers is relatively small, namely mothers that are older than 35 when giving birth for the first time. All of the result confirm our main hypothesis that traditional estimates using twinning are bound to be downward biased.

The third paper, entitled "Child Gender and its Effects on Parental Labor Market Participation: A Robust Tale of Danish Parents" is also concerned with the effect that children have on parents in the years after birth. However in this case we look at whether having a boy or a girl respectively, tends to significantly affect parental behavior. Again, there is a vast economic and sociological literature analyzing the effects of child gender on such diverse outcomes as parental time use, political orientation or fertility behavior. Most studies do however simply treat child gender as being random and exogenous. In biology and demography there is a sometimes controversial strain of literature arguing that this is not necessarily the case. A series of studies have shown that ethnic factors such as race, biological factors such as age and even economic stress can affect the sex ratio, or relative quantity of male to female children that are born to mothers. Most studies find results pointing in the direction that better parental health and less stress tend to be related to a slightly higher relative prevalence of boys. This opens up the possibility that some of the results we encounter for child gender are driven by omitted variable bias.

We thus analyze exactly which factors influence the sex-ratio at birth in the Danish population from 1980-92. We then go on to see whether inclusion or exclusion of the factors that we find to have an effect in our regression on birth probabilities, has any mediating effects when looking at the effect of child gender on parental behavior. Maternal education and paternal age, as well as birth year are found to have significant effects on the probability of having boys. We then go on to estimate the effects of child gender on maternal and paternal income and employment. When checking whether any of these factors tend to have mediating results we find only very minor, insignificant changes in the coefficients. This strengthens the case that most effects that the literature on the consequences of child gender finds are genuine. In general we find fewer effects of child gender on parental outcomes than most of the literature established. Neither maternal labor market outcomes, nor divorce or marriage rates are affected by the gender of the child. However men tend to have lower incomes when they have boys than when they have girls. In contrast to other studies finding effects shortly after birth we find that the negative effect of having a boy on parental earnings tend to arrive relatively late (about 8 years after birth) and to increase with time. We also find that these effects are stronger for highly educated couples. All of these results are highly stable to specifications aiming to control for factors influencing sex-ratios.

Overall I conclude from this thesis that when using randomization techniques for causal identification it is worthwhile to follow Leamer's advice and to not solely rely on study design but to instead check for sensitivity and to think about causal identification techniques more deeply embedded in the context of what is being studied. Three particular questions that are worthwhile asking and that have each been respectively addressed by one of the papers are :

1. Does the standard estimator used for evaluating experimental studies, namely the Local Average Treatment Effect (LATE), which measures the average difference between a group that has been randomly exposed (or treated) to whatever policy or event we wish to evaluate, really capture the important aspects of what exposure to the policy or event means? Might it be worthwhile exploring whether the impact differs across the distribution? This is particularly worthwhile if there are good theoretical reasons to assume heterogeneous treatment effects across the distribution of individuals. The paper on non-linear

preschool effects explores this aspect.

2. Might there be unexpected consequences of the random variation used for identification, that can influence the outcome variable in a way that is not due to the causal relationship, which is being explored? The second paper exploring how subsequent fertility behavior can bias estimates explores this aspect.

3. Might there be reasons for which the variation, which is being used might potentially not be entirely random? Can we then find ways to assess whether factors impairing complete randomness matter for the estimates? The third paper looking at whether selection affects our estimates of child gender on parental outcomes explores this aspect.

As we will see a cautious approach to experimental or quasi-experimental evaluation can both overturn or confirm previous results. Most importantly, however, in either case it tends to add nuance and insight to our results.

2 A non-linear Assessment of Preschool Effects

Abstract

There is an apparent contradiction between recent studies on early childhood interventions, which find significant and large gains on a variety of outcomes for the treated children and a series of studies on the effects of universal childcare, which tend to find small and often insignificant effects. A recently proposed explanation (Havnes and Mogstad 2010) is that the lower effects in universal childcare are a result of the different sample composition and that universal childcare has highly beneficial results on more disadvantaged individuals, but that much of this is lost in mean treatment estimations. We revisit two studies on the effects of preschool in Denmark, which use an instrumental variable approach to estimate effects of preschool on cognitive and non cognitive child outcomes at age 10. In general the non-linear estimates, obtained by using instrumental variable quantile regression confirm the initial results of the Danish studies. However some new results do appear. Preschool has the biggest positive effects on cognitive and non-cognitive outcomes of poorly performing boys. Surprisingly preschool affects non-cognitive outcomes for girls in the lowest quantiles in a significantly negative way.

2.1 Introduction

In the quest of social scientists to uncover the forces that shape our long-term socio-economic outcomes, there is a sense that the further back we look, the more profoundly the effects we uncover tend to shape us. Such is the idea behind Heckman's returns on learning curve (Heckman et al. 2006). It theorizes that the earlier an aspect of learning occurs, the bigger lifetime expected returns in terms of socio-economic performance tend to be. This has profound policy implications and thus demands rigorous testing. This paper is a contribution to the growing literature on returns to Preschool Education. While most of the literature agrees, that effects of early childhood interventions are substantial the few evaluations that exist of universal childcare tend to show weak or non-existent effects. (Datta Gupta and Simonsen 2011) (Havnes and Mogstad 2011) (Lefebvre et al. 2008). There are different possible explanations for this apparent contradiction. The first explanation would be that small-scale targeted interventions are simply more effective than universal early education systems. This might be due to their more specific and targeted design or due to problems in successfully scaling such programs. Another explanation would be that we are better able to evaluate small-scale interventions, since they often employ experimental designs, including control groups and thus lend themselves much better to causal evaluation. This hypothesis assumes that for methodological reasons we simply are not able to find the positive effects of universal child care programs. A third explanation argues that sample selection might be the main driver of the differing results. Targeted programs generally target disadvantaged children, if those children benefit substantially more from early education interventions, this would explain why effects found for universal programs tend to be so much smaller. Recent work on the introduction of universal childcare in Norway (Havnes and Mogstad 2010) provides evidence that is strongly in line with the third explanation. Using variation in the introduction of childcare to different areas of Norway in order to estimate difference-in-difference effects the Norwegian study initially find little impacts of universal childcare. However, applying non-linear estimators it concludes that the effects are highly non-linear across the distribution of children with those coming from more disadvantaged situations, being affected quite strongly, an effect that is lost in mean estimates. This paper aims to provide further testing of the idea that universal childcare systems tend to provide strongly non-linear treatment effects, mainly lifting those at the bottom of the distribution up. Policy initiatives promoting the expansion of child-care, such as the 'Barcelona targets' of the EU commission, which promote the expansion of early childcare are increasingly based on the implicit assumption that such initiatives would in particular help lower performing children. It is thus important to rigorously test these ideas.

We use high-quality data from the Danish Longitudinal Survey of Children (DALSC) which includes extensive information on the family and schooling situation of children as well as cognitive and non-cognitive test-scores. The data on parents was merged with the Danish registry data in order to obtain well-documented labour market information on the parents. We look at the effects of attendance of different types of preschool programs on results in reading and mathematical reasoning test scores as well as on a series of behavioral scores at later ages, all of which are recorded in the DALSC data. To do so we exploit a unique setup resulting in quasi-experimental variation in the supply of high-quality preschool programs in Denmark. The policy setup as well as the data have previously been exploited by Simonsen and Gupta (Datta Gupta and Simonsen 2010) . Their work suggests that Danish high-quality preschool compared to lower quality preschool at age 3 shows no discernible effects on cognitive test scores at age 7. In a separate study a number of weak effects on risky behavior at age 11 are found when comparing preschool to family day-care, but again no effects show up when comparing high-quality to low quality preschool (Datta Gupta and Simonsen 2011).
This paper expands on those results by providing a series of non-linear assessments on the cognitive and non-cognitive test-score measures taken at age 7. As we outlined choosing a non-linear approach seems logically consistent with the fact that the strong results of studies on targeted interventions might be due to them treating a more at-risk-population. It might thus be that stronger effects at the bottom of the distribution, simply became diluted by the inclusion of kids from better socio-economic backgrounds into the treated population as was the case in Norway. Estimation methods that split the data into subgroups are only a very unsatisfactory solution in this case, as the sample size is already relatively small and as intergroup variance often tends to be substantive. This variance would be completely lost by separately calculating regressions for subgroups, which only account for within-group variation in the data(Havnes and Mogstad 2010). By addressing these problems with several newly available non-linear estimators our aim is not only to contribute to the literature on

the effects of preschool programs but also to highlight a series of important new statistical tools for policy evaluation, where outcomes are expected to be non-linear.

We further decided to estimate the nonlinear treatment effects separately for boy and for girls. Not only have female enrollment rates in university surpassed those of men in many countries but there is also a literature stressing that particularly at the lower ends of the performance distribution, it is increasingly boys that seem to particularly struggle in school. For a review see Buchmann et al. (2008). Recent findings have linked much of the higher female college enrollments to their higher non-cognitive skills after high school Jacob (2002). We thus think it is important to look at what might influence these gender differences in non-cognitive skills at early stages. In particular due to the findings regarding "troubled boys" we expect that better preschool care might have particularly beneficent effects on boys in the lower end of the distribution.

Our outcome variables include a mathematical reasoning score a reading score and the SDQ (Strength and Difficulties) test for behavioral outcomes. SDQ has been shown to provide more consistent and broader screening results for children's mental health than the previously common Rutters test (Goodman 1997). The importance of assessing the development of non-cognitive skills has been highlighted by work showing that early development of non-cognitive skills has a significant impact on schooling decisions and subsequently continues to influence wages, given schooling decisions (Heckman et al. 2006).

Applying the outlined previous findings to our study we thus have three main hypotheses.

1. We expect the effects of preschool on cognitive and non-cognitive skills to be non-linear and more positive in the lower ends of the distribution.

2. The effects are likely to be more positive for boys than for girls.

3. Given that we compare high-quality preschool to lower quality preschool (instead of to no preschool) our counterfactual does not capture such a significant difference as in other studies. Thus effects might be relatively small.

We find results that are broadly consistent with the original linearly obtained results of the previous study done with this data (Datta Gupta and Simonsen 2011). However some non-linearities do show up. More particularly we find that under performing boys benefit most in terms of cognitive and behavioral scores from attending preschool instead of family day-care programs. Surprisingly we find that preschool has significant negative effects on the behavioral scores, of more at-risk girls.

2.2 Child Care

A simple argument for investing into children at a young age goes back to work by Becker. It stresses that earlier investments simply allow the recipients to reap the benefits for a longer period of time (Becker 1994) than later ones. More recently a cross-disciplinary consensus on the benefits of early childcare, which goes far beyond Becker's initial argument has emerged. Investments at the early stages of life might result in additional benefits for a variety of reasons. neuroscientists have shown the extent to which the early childhood period is crucial to the formation of neural networks, mediating sensory, motoric, social,linguistic and emotional capacities (Katz and Shatz 1996) (Knudsen 2004). Psychologists have shown how important, both positive and negative experiences in early life are for the formation of social competences. Further, economists and sociologists have shown, how important experiences in early life are for the formation of human capital and for assuring future economic success. The risk factors that identify children as disadvantaged have also broadened in scope. Clearly the most heavily studied "risk factor" has been poverty. But parental education, mental health problems and exposure to conflict and violence, have also been studied recently (Shonkoff and Phillips 2000). Knudsen et al (Knudsen et al. 2006) provide a synthesis that particularly highlights how important investments into disadvantaged children are. In this context Currie stresses that governments should aim to equalize life-chances early through investments into childcare instead of aiming to compensate for those differences later in life (Currie 2001). Evaluations of early childcare in the US have mostly found positive results on children's outcomes. For example Fitzpatrick finds strong positive effects of Pre-K on later test scores (D 2008). In another US study Loeb et al found positive effects on reading, math and behavioral score measures (Loeb et al. 2007), whereas a follow up study using the same data found negative behavioral effects and non-lasting cognitive improvements (Magnuson et al. 2007). Positive effects on later school performance were also found when looking at pre-primary education in Uruguay (Berlinski et al. 2008). One of the few studies finding negative effects was that of Herbst and Termin which looked at the effects of child-care subsidies given to welfare recipients in the United States (Herbst and Tekin 2010).

Yet, while the importance and effectiveness of interventions aimed at disadvantaged children has been shown to a great extent, the literature on the effects of universal childcare remains relatively limited. The effects of universal child-care cannot necessarily be expected to be the same as those of specialized interventions, as universal programs are not specifically targeted at children at risk. The aims of universal childcare are also different, often aimed at allowing parents to combine having children with pursuing a carreer and are not necessarily in the same way targeted at improving the outcomes of a particular group of children.
So far there are three universal child-care programs that have been evaluated in a setup aiming to derive causal estimates of the outcomes achieved by such systems. Those are Quebec, Norway and Denmark.
In the late 1990s Quebec introduced a universal, highly-subsidized childcare system. A careful analysis of the before and after variation in outcomes in Quebec and surrounding regions which did not introduce such childcare systems led Baker et al. to conclude that the introduction of childcare had strongly significant positive effects on female labor supply but that effects on cognitive outcomes as well as non-cognitive behavioral measures were significantly negative, while indicators of family stress went up (Baker et al. 2008).
The second case which has been intensely studied is Norway, which started significantly expanding childcare availability in the late 1970s, slowly building up towards a system of universal care. Havnes and Mogstad exploited the regional variation in the speed of the introduction of child-care to obtain estimates via a difference-in-difference approach for Norway (Havnes and Mogstad 2011). In general they found small and insignificant positive effects on later earnings of children exposed to childcare. However they reestimated their results using non-linear difference in difference methods and in that case found substantial positive effects in the lower range of the earnings distribution, making the point that mean estimates might miss the most important aspects of the effects of universal childcare. In the case of Norway Rindfuss et al. also explored the

effects of childcare on fertility decisions of parents and found significant positive effects Rindfuss et al. (2007).
Finally, as discussed, evaluations of the Danish child-care system, show no effects of preschool and negative effects of daycare on non-cognitive skills at age 7 (Datta Gupta and Simonsen 2010), further no effect of the Danish childcare system on cognitive skills could be found (Datta Gupta and Simonsen 2011). It has to be noted however that in the Danish case the counterfactual is relatively weaker than in the Quebuecois and Norwegian studies. While those studies look at the introduction of an all-together novel universal childcare system the Danish study looks at assignment to different childcare programs with differing quality. Estimates thus do not reflect the effect of receiving childcare as opposed to not receiving it, but rather the effects of attending different types of programs.

Nevertheless the question arises whether in the Danish case high-quality preschool also tends to lift up more disadvantaged children as was the case in Norway and in targeted studies. A key point of Havnes and Mogstad's Norwegian study is that it criticizes subdividing the data along certain risk criteria, since between group variation is lost. Estimating with a non-linear difference in difference estimator they find that non-linear treatment effects exist, which they were unable to uncover with mean estimates. This paper expands on the idea that non-linear estimators might be much more adept to capture effects on individuals across the distribution. Instead of a difference in difference approach we revisit the studies on Danish universal childcare using non-linear instrumental variable estimators to see whether effects tend to differ strongly across the distribution in Denmark as was the case in Norway.

2.3 Background and Data

2.3.1 Childcare in Denmark

The increasing labor-market participation of women has been accompanied by an increased demand for and supply of early child-care programs across OECD countries (Jaumotte 2003a). Since there is a broad general trend towards more early childcare, Denmark with its already extremely well-built child-care support offers in many ways a perfect case study of the impacts these programs might have in other countries that move towards their implementation. From early on Denmark has been a

forerunner in increasing access to early childcare, having implemented a system of universal high-quality care in the 1970s. In 2007 Danish Public spending on childcare was at 1.4% of GDP, more than in any other OECD country (OECD 2011).
Until age 2 the state generally provides parents with nurseries. After that there is three options for continued childcare: preschool, family day care and home care. What follows is a brief summary of the types of care, offered at age 3 to children and of their allocation. This has been outlined in detail in previous work by Simonsen and Gupta. (Datta Gupta and Simonsen 2010) (Simonsen 2010)
Preschools have high-standards in the quality norms they enforce. There are on average about 9 educators for 60 children, which are split into groups of 20. The educators are supplemented by additional staff, freeing them up for educational activities. A high degree of job preparation is ensured by the requirement that preschool teachers in permanent positions must have a degree in teaching that specializes in young children. Preschools are generally open from 6.30 a.m. to 5 p.m.. However municipalities may vary these hours as long as they ensure that "local needs" are covered. It is also in the charge of municipalities to ensure that hygiene standards, access to play facilities and safety measures are in place.
Family Day Care is provided by caretakers, taking up to 5 children in their house. This may include their own children. The municipality then pays them for taking care of the children. The educational requirements for educators are less strict than at preschools. A degree in teaching is not required, instead there are three-week vocational courses being offered for caretakers.
Home Care A final option is that the parents simply decide to take care of the children at home.

The universal availability of child-care is ensured via a generous system of subsidies. Families making under $ 20.000 a year will receive the child-care for free. Families making under $ 60.000 a year will receive an income-dependent subsidy. The maximum price for family day-care is set at $ 3.500 a year and for preschool at $ 2.600 a year. This covers about 33% of the actual cost.

2.3.2 Allocation of Childcare and Guaranteed Access to Preschool

An important characteristic of the Danish child-care system is that it is not allowed to exclude any children for social or economic reasons from the access to child-care. The parents have to apply to child-care slots directly to the municipality. This ensures that relationships of parents to the caretaking institutions cannot influence the selection process.
Depending on the municipality, a waiting-list, which is either ordered by the child's birth-day or by the date of application is put together. Parents can indicate whether they prefer preschool or family day-care. Then slots are allocated according to the waiting list. If the parents are not content with the slot they have been allocated, they can demand that their child is put on the waiting list again on a later slot until they receive a new offer. This means that parental patience as well as the strength of parental preference, which might be related to other characteristics can influence the child-care that a given kid ends up in.
There are very few characteristics which allow the child to jump ahead in the waiting list. These are being disabled, being an immigrant, or having older siblings enrolled in municipality provided care. We exclude immigrants from our estimations, because of problems in the DALSC sample collection of immigrants. This is regrettable since, in particular many of the distributional and of the gender differences in educational attainment that we are looking for, have been documented to be particularly strong among immigrants Lopez (2003). Further we include controls for the child's health status and for the presence of older siblings, since this might influence the selection process.

The municipalities are required to provide sufficient child-care slots, so that all children will be taken care off. It is left to the municipalities to decide on the ratio between pre-school and family day care slots they provide. Some municipalities do however have a program called Guaranteed Access to Preschool (GAPS) in place in certain years, which as we will see leads to higher take-up rates of the preschool option. This strongly supports the logical notion that the pre-school option is indeed preferred by a vast majority of parents, which is crucial to our identification strategy. Simonsen and Gupta (Simonsen 2010) argue that the provision of GAPS is most likely to stem from random shocks in cohort size. Their argument is fairly simple. Providing too little preschool will lead to voter

Variable	Preschool	Daycare	Homecare
Overall	.736	.186	.078
GAPS municipalities	.903	.048	.049
Non GAPS municipalities	.648	.259	.094

Table 2.1: Types of Childcare in GAPS and Non GAPS municipalities

dissatisfaction, whereas providing too many slots imposes unnecessary costs on the municipalities. Since construction of new preschools imposes relatively high costs, it is hard to adapt the offer to short-term variations in cohort-size. Thus changes in cohort size will determine the municipalities capacity to provide GAPS or not. This leads to a relatively unpredictable variation in GAPS treatment. Given that these variations can easily change the GAPS status of a municipality from one year to another it also makes it more unlikely that families decide where to settle in order to take advantage of child-care policies. Simonsen and Gupta also show that living in a given municipality is not influenced, much by the set of controlling variables. GAPS thus induces variation in the take-up rates of preschool which is not influenced by parental decision-making.

Since parents face 3 potential options we find ourselves in a case of treatment with multiple outcomes. For our estimations it is important to assume that there is a weak preference for preschool. More precisely, the GAPS treatment should induce some parents, who in other communities would have ended up sending their kids to daycare or to taking care of them at home, to send their kids to preschool. The high uptake of preschool in GAPS relative to non-GAPS municipalities provides good evidence that this is indeed the case, see Table 2.1. It has to be noted however that uptake of high-quality preschool relative to daycare is the case in which our identification strategy clearly works best as there is a well defined preference and parents who send their kids to daycare can uniformly be assumed to have sent their children to preschool had the option been available to them. The same does not apply for home care, which does not depend on the slot allocation of the municipality and is thus much less clearly treated by the GAPS variation, even though we find that availability of the GAPS program is also associated with lower numbers of children in homecare.

Data

The data we use is from the Danish Longitudinal Survey of Children (DALSC). It is a longitudinal survey of children born between September 15th and October 31st 1995. The closeness of births of the children included in the sample eliminates the possibility of significant cohort effects. The child data is merged with high-quality registry data, providing background on parental education, labor market status and income over the same period that the DALSC data covers. Currently 4 waves of data on the children are available, they were collected in 1996, 1999, 2003 and 2007. A fifth wave covering 2010 is about to become available. The information on the type of child-care enrollment is taken from the 1999 wave, when the children were 3 1/2 years old. As outcome measures we use a behavioral score , which is based on parents filling out questionnaires, as well as on reading and mathematical reasoning test scores of the children. The SDQ (Strength and Difficulties Questionnaire), which is our behavioral score measure was collected twice, once in 2003 and once in 2007. The Children's Problem Solving Test (CHIPS) and reading ability test-scores were collected in the 2007 wave or at age 11 1/2.

2.3.3 Variables of Interest

We are interested in child outcomes and the influence the childcare system has on them. We have a rich variety of different measures of child outcomes. Our first measure is the Strength and Difficulties Questionnaire (SDQ), which we have available for two different years, 2003 and 2007. SDQ is obtained by asking parents 25 questions on their child's emotional symptoms, conduct problems, hyperactivity/inattention problems, peer relationship problems and pro-social behavior. These 5 topics are called subscales, which are each made up of 5 questions. The answers of parents are scored from 0 to 2 (with a higher value indicating more problematic behavior. The SDQ thus ranges from a score of 0 to 50 (with 0 indicating the best possible behavior).

Despite being relatively new SDQ has already found widespread acceptance as a behavioral measure (Goodman 1997). It correlates highly with the otherwise used Rutters Scale for evaluating child behavioral problems. [1]

[1] The full set of questions can be visited at www.sdqinfo.com.

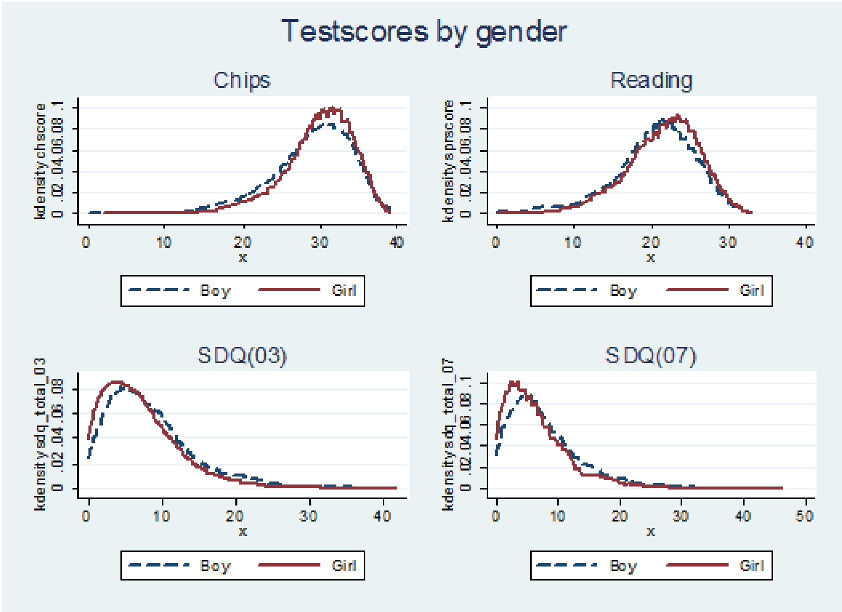

Figure 2.1: Distribution of Testscores by gender

The other outcome measures we use are a language test and the Children's Problem Solving Test (CHIPS), which measures cognitive reasoning and a reading test. Both tests consist of a series of multiple choice questions, 40 in the case of the CHIPS test and 34 for the reading test. The scores thus range from 0 to 40 (0 to 34) with a higher score indicating a better performance.

Figure 2.1 shows the distribution of test scores by gender. We can see that on every measure girls perform better. On the Chips and reading tests, where higher scores indicate a better performance, their distribution is more to the right and on the SDQ measures, were lower scores indicate a better performance, their distribution is to the left of the boys.

Our models include a set of child and parental characteristics as controls. Since handicapped children might be treated preferentially in gaining access to preschools and since handicaps might also affect test outcomes, we include a control for handicaps at birth in our set of controls. We also include birth weight as an indicator of health, and whether the child was breastfed, as various studies attest for its effects on child outcomes. We also include a binary variable on whether the kids were born in a rural area.

Variable	Pre-School	Family Day Care	Homecare	Total
Child Characteristics				
birth handicaps (96)	.04	.04	.02	.04
birthweight(96)	3508	3528	3559	3507
breastfed (96)	.96	.95	.95	.96
siblings (96)	1.77	1.81	2.27	1.83
SDQ (03)	7.76	7.98	7.38	7.82
SDQ (07)	7.00	7.61	6.70	7.15
SPROEG score (07)	21.24	20.23	20.629	20.96
CHIPS score (07)	29.24	28.73	29.20	29.08
siblings (07)	2.19	2.22	2.28	2.25
health score (07)	1.62	1.61	1.64	1.62
Mother Characteristics				
Age at birth (96)	29.4	29.3	29.4	29.3
ISCED (99)	3.63	3.38	2.85	3.52
Father Characteristics				
ISCED (99)	4.43	3.94	4.04	4.31
Family Characteristics				
rural(96)	.12	.26	.35	.16
log adjusted family income (03)	12.13	12.04	11.94	12.01
reading to child (03)	.80	.76	.74	.79
divorce (07)	.24	.12	.19	.23

Table 2.2: Summary statistics by type of childcare

Having older siblings in daycare programs is also a reason for potential preferential treatment which is why we include the number of siblings at birth. We further include the number of siblings, at the age that tests are taken in order to fully account for all sibling effects. At the age of test-taking we also include a health score for the children. This is based on an assessment of the mother in 2007, in which she describes the health of her child as ranging from very good(1) too bad(4). We also record whether the biological parents divorced prior to 2007.
For both parents we include the level of education on an ISCED 8 scale. As a proxy for parental interaction with the child we included a variable on whether the parents regularly read to the child at age 7. Finally we calculated log adjusted family income by logging the joint income of both partners (if they are jointly taking care of the child) and then dividing it by the square root of the number of family members. Table 2.2 provides a summary of the variables by type of childcare. Note that in the summary table the year of the survey is noted in brackets behind the variable name. Be aware of the fact that certain control variables such as reading to child or divorce have been recorded post-treatment. Meaning that it cannot be entirely ruled out that they were contaminated by the treatment even though we consider these effects to be most likely very minor [2]

2.4 Empirical Strategy

We present the results on three different outcomes, SDQ (07), CHIPS and Reading Tests [3]. To make regression coefficients as comparable as possible across our different outcome measures we standard-normalized each outcome and inverted the values for the SDQ test, so that higher values now correspond to better outcomes. Since we estimate each of our models on the 3 2007 test scores we will in the following abbreviate the joint set of our dependent variables with TS, for test score.
As we discussed previously, when looking at childcare in Denmark we are placed in an environment with multiple treatment outcomes. Theoretically we might be interested in three possible outcomes. The first one of these is:

[2] Also note that our key findings, including the nonlinear ones were fairly stable to the inclusion or exclusion of the control variables.

[3] SDG(03) had results very similar to SDQ(07) but a less rich set of available controls. Therefore we did not explicitly report these regressions

$$E(TS_{FDC} - TS_{HC}|FDC = 1) \tag{2.1}$$

This equation captures the difference between the expected value of test scores for those who participate in family daycare and those who participate in home-care, for the treated. For the treated, means that the treatment condition applies or that values are conditional on individuals actually switching to family daycare.
The second outcome we could consider is

$$E(TS_{PS} - TS_{HC}|PS = 1) \tag{2.2}$$

This captures the same difference but with preschool instead of family daycare as a counterfactual to homecare.
Finally we consider:

$$E(TS_{PS} - TS_{FDC}|PS = 1) \tag{2.3}$$

This denotes the expected test score of preschool relative to family day-care, for children enrolled in preschool.

All three equations give us a parameter capturing the difference of being exposed to one type of care instead of being exposed to another, conditional on switching exposure due to treatment. Estimating the first two equations would however suffer from the fact that parents may not randomly select in and out of home-care. If better quality preschool becomes available and some parents therefore decide to send their children into preschool instead of taking care of them at home, while some do not, it is rather likely that this decision is related to lots of unobservable parental characteristics that we cannot properly asses. Therefore the only counter-factual we are able to properly estimate is the one in equation 2.3. Parents that decided to enlist their children for publicly provided care but switch from family daycare to preschool are not affected by these selection problems. Our models and estimations will thus only aim to capture the treatment effect described in the last equation. The treatment effect which we assess via the instrumental variable setup can thus be thought of as follows:

$$E(TS_{PS} - TS_{FDC}|PS(GAPS) - PS(NOGAPS) = 1) \tag{2.4}$$

This means we capture the difference in the test scores of those going to preschool TS_{PS} and those going to Family Daycare TS_{FDC} for those individuals that go to preschool if there is a GAPS treatment but would not have gone to preschool if there was no GAPS treatment, which is captured by the expression $PS(GAPS) - PS(NOGAPS) = 1$.

In the following we will first estimate our model comparing preschool to family daycare via OLS. Then we will look at a quantile regression version of that model. Next we reestimate the OLS model using GAPS as an instrument for preschool. Finally we estimate a quantile instrumental variable versions of our model. We use the conditional instrumental variable estimator for quantile treatment effects, proposed by Abadie,Angrist and Imbens (Abadie et al. 2002). We also implemented the unconditional IV estimator proposed by Froehlich and Melly (Frölich and Melly 2008). The difference between these estimators mainly comes down to the fact that if using the unconditional IV estimator the covariates do not influence our treatment parameter estimates.
The quantile treatment estimator requires a binary treatment indicator and a binary instrument, both conditions which are fulfilled in our case, with participation in preschool as well as exposure to GAPS being binary. Further it requires that the four instrumental variable assumptions are fulfilled. These are

1. Independence: as discussed in detail earlier we do not expect that the availability of GAPS is in any way related to variables influencing our outcomes

2. Non-trivial assignment: this simply means that the share of treated or non-treated individuals is not equal to 0 or 1 as is the case.

3. First Stage: this means that treatment, or availability of GAPS, does actually lead to higher preschool enrollment rates as shown [4]

4. Monotonicity: This means that the availability of GAPS leads some parents to switch from family daycare to preschool but does not induce switches in the other direction. This is impossible to prove, since we cannot construct counterfactuals at the individual level but it is a very reasonable assumption.

[4] for reasons of parsimony we did not include the first stage regressions in the paper. They are highly significant and basically reflect the effect of GAPS treatment that table 2.1 captures. They are available upon request.

The exact derivation of the quantile treatment estimator we use is described in great detail in the work of Abadie,Angrist and Imbens (Abadie et al. 2002). The key parameter of interest they derive is α_θ which in our case gives us the difference in test scores between treated (exposed to GAPS) and non-treated individuals at the $\theta - quantile$ for compliers.

2.4.1 Linear Model

The results presented for our OLS estimation compare preschool to family day care. Individuals, who attended home-care were excluded from the sample. Results are presented for our three outcome measures. You can see from Table 2.3 that attending preschool has a positive coefficient in each regression. In those regressions that were done for the entire sample of boys and girls, the effects are significant, for the reading scores and for the cognitive tests. Obviously these results can only be viewed as a benchmark since selection is prone to influence the results. Parental education and income, as well as the gender of the child are the controls which most significantly affect outcomes. Interestingly divorce has strong effects on non-cognitive skills, while the effects on cognitive skills remain insignificant.

2.4.2 Non-Linear Model

As previously discussed we suspect that the effects, preschool treatment has on children might be highly non-linear. To assess this we run the entire set of OLS models displayed in table 2.3 as quantile regressions. These estimates are similar to what has been done in previous work by Esping-Andersen et al (Esping-Andersen et al. 2012). Fig.2.2 shows how the regression coefficient for preschool develops across quantiles. In particular for the CHIPS and the reading score it seems like effects are highly non-linear and strongly positive in the lowest quantiles. The 5% confidence interval is included in gray around the estimates to allow assessing the significance of the estimates. The mean estimate and its confidence interval are included as dotted line. We can see that in the quantile estimates we continue to encounter meaningful positive effects of preschool (as opposed to daycare) on the test-scores. Our first hypothesis, on finding stronger positive effects at the bottom of the distribution seems to hold op for the reading and CHIPS reasoning scores, but do however not show up in our estimates on the behavioral SDQ scores.

Variable	Reading			Chips			SDQ		
	Total	Boys	Girls	Total	Boys	Girls	Total	Boys	Girls
schooling									
preschool (03)	.108*	.121	.087	.101*	.128	.063	.078	.139	.016
child char.									
Boy	-.235***	n.a.	n.a.	-.249***	n.a.	n.a.	-.226***	n.a.	n.a.
birth handicaps (96)	-.033	-.130	.115	.145*	.084	.251*	-.005	-.019	-.008
birthweight(96)	.001**	.001**	.000	.001***	.000	.001***	.000	.000	.000
breastfed (96)	.231	.265	.214	.208	.360	.079	.199	.123	.260
siblings(96)	-.068*	-.011	-.134***	-.033	-.022	-.052	.013	-.027	.061
siblings (07)	.058*	.070	.057	.051	.020	.081*	.063	.094	.032
health score (07)	.027	.079	-.030	-.007	.017	-.027	-.124**	-.132*	-.117
mother char.									
Age at birth (96)	.009	.007	.013	.001	.005	.005	.013*	.023**	.002
ISCED (99)	.033**	.023	.039**	.007	.003	.013	.026*	.019	.034*
father char.									
ISCED (99)	.068***	.061***	.076***	.050***	.059***	.040*	.015	.012	.017
rural(96)	.044	.107	-.003	.074	.061	.021	.023	.116	.012
log adj. family income (03)	.311***	.463***	.134	.188**	.255**	.097	.203***	.222*	.169*
reading to child (03)	.041	.001	.084	.050	.091	-.003	.076	.032	.124
divorce (07)	-.029	-.029	-.021	.013	-.005	.025	-.261***	-.227*	-.298***
N	2562	1330	1232	2582	1349	1233	2564	1339	1225
R^2	.069	.073	.087	.049	.047	.027	.066	.054	.058

Table 2.3: OLS estimations for preschool and family daycare

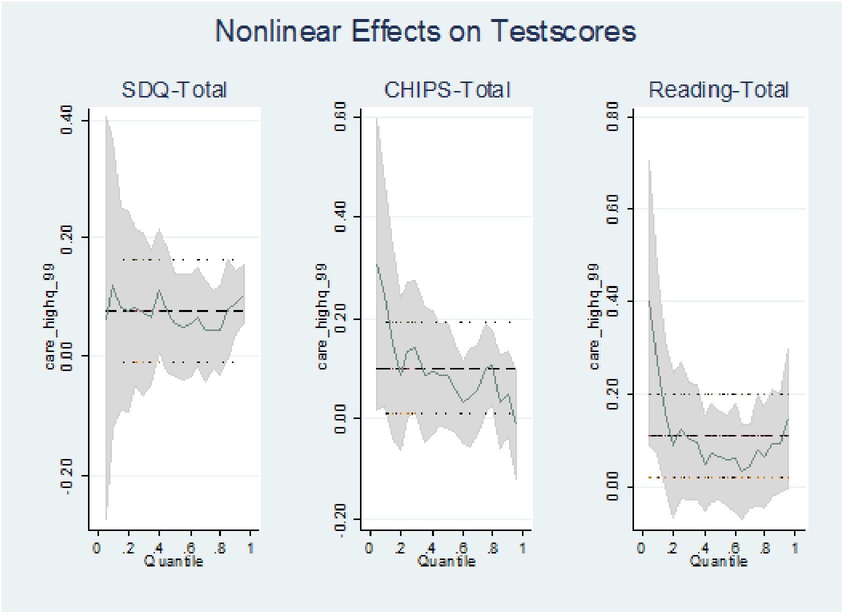

Figure 2.2: Quantile Regression, Coefficients for Preschool on Testscores

As discussed we also decided to run the quantile regressions separately for boys and girls, since there is much evidence showing that boys are particularly responsive to the effects of positive as well as negative environments. We thus concluded that to fully account for the nature of non-linear effects it might be necessary to look at the effects separately by gender. Figure 2.3 shows those results. For boys we consistently see that effects are strongest in the lowest quintiles, while for girls this is not the case. Particularly striking, is how the effects of preschool on our non-cognitive skills measure differ. For the lowest performing girls, preschool seems to actually worsen the outcomes measured by SDQ, while for boys the opposite holds true. The linearity on SDQ we found in our estimates using the entire sample thus masked two opposing types of non-linearity for boys and for girls.

2.4.3 Instrumental Variable Regression

Surprisingly, using instrumental variable estimation does not only lead to weaker coefficients, as is often the case, but even the signs in front of some coefficients change, even though not in a significant way. In particular

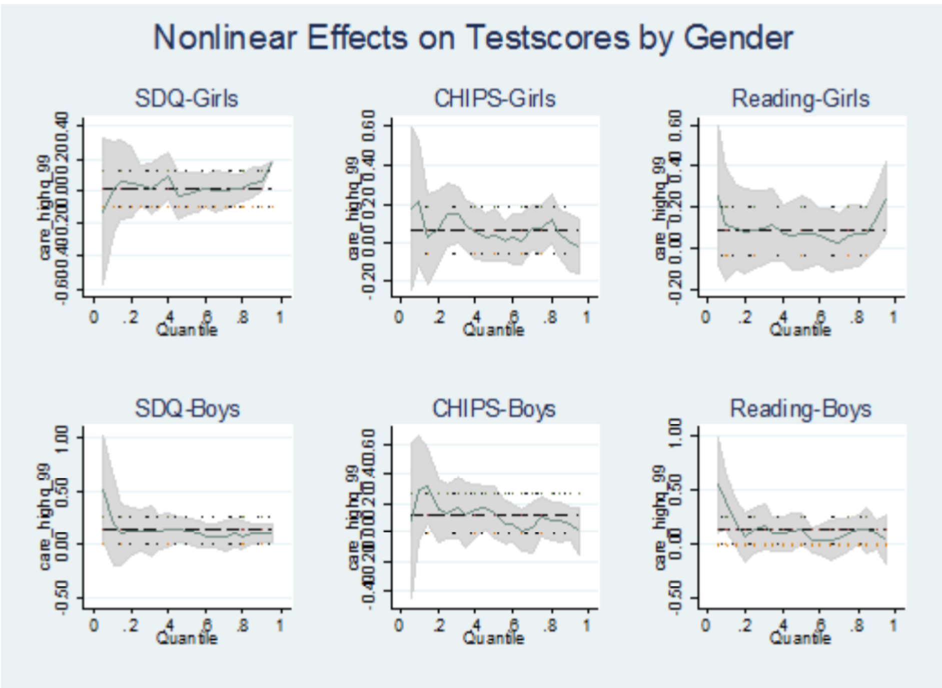

Figure 2.3: Quantile Regression by Gender, Coefficients for Preschool on Testscores

we now have negative effects on the overall reading score, as well as on the SDQ of girls. Again results differ strongly by gender. For reasons of parsimony the first stage regression is not included, but it turns out highly significant and the descriptive statistics in table 2.1 already show the extent to which GAPS treatment led to higher enrollment in preschool. The full instrumental variable regression results can be seen in table 2.4.

2.4.4 Non-Linear Instrumental Variables

To obtain non-linear IV estimates, we used both conditional and unconditional estimators. In unconditional estimators the coefficient for the treatment variables are not influenced by the covariates. Using both estimators served as a robustness check and indeed the results are very similar. Below we report only the results from the conditional estimator. In Fig. 2.4 we can see that while some non-linearities do appear, the results overall are consistently close to the original findings of Simonsen and Gupta, who found only small or non-existent effects of preschool (Datta Gupta and Simonsen 2010). None of our linear IV estimates shows a treatment effect that is actually bigger than one standard deviation of the test score. We also find some, but not overwhelming evidence for our first hypothesis of nonlinearities. The findings are also more contradictory than expected, with Chips scores displaying some of the positive nonlinear effects we were expecting at the bottom of the distribution, but SDQ instead displaying negative effects, in particular for lower performing girls. We do thus find confirmation for our second hypothesis that we expect boys to be more positively affected than girls. Indeed the negative effect of preschool on the low-scoring SDQ kids seems to be almost entirely driven by girls and the positive effects at the lower end of Chips Scores are mainly driven by boys. We thus find some confirmation for what we expected, but also, in the female SDQ scores some relatively unexpected results. When trying to make sense of these findings it is important to keep in mind the exact counterfactual that these estimates compare. We are after all not comparing preschool to homecare, but instead preschool to family daycare. This means we are comparing two different programs of family daycare in Denmark, where the quality of such programs is generally high. Thus the variation between the treated and the treated can be expected to be relatively smaller than in many other countries. Family daycare is generally considered to be of lower quality but might also have some aspects where it is hard to predict whether they turn out advantageous or disadvantageous for a specific child, such as a very

Variable	Reading			Chips			SDQ		
	Total	Boys	Girls	Total	Boys	Girls	Total	Boys	Girls
schooling									
preschool (03)	-.090 *	-.259	.066	.146	.070	.063	-.135	.322	-.649*
child characteristics									
Boy	-.232***	n.a.	n.a.	-.250***	n.a.	n.a.	-.224***	n.a.	n.a.
birth handicaps (96)	-.030	-.131	.115	.146*	.085	.251*	-.006	-.022	-.026
birthweight(96)	.001*	.001*	.000	.001***	.001***	.000	.000	.000	.000
breastfed (96)	.237*	.276	.215	.206	.356	.079	.205	.117	.282
siblings(96)	-.066*	-.006	-.134***	-.033	-.024	-.051	.016	-.031	.064
siblings (07)	.058	.067	.057	.052	.020	.081	.063	.095	.039
health score (07)	.029	.091	-.030	-.008	.014	-.026	-.121**	-.137*	-.118*
mother characteristics									
Age at birth (96)	.008	.006	.013	.001	.005	.005	.013*	.024**	.002
ISCED (99)	.032**	.018	.040**	.008	.003	.013	.024*	.022	.037*
father characteristics									
ISCED (99)	.072***	.072***	.076***	.049***	.056***	.039*	.020	.007	.024
rural(96)	.016	.059	-.007	.080	.144	.021	.036	.138	-.077
log adj. family income (03)	.341***	.535***	.136	.181*	.237*	.097	.233***	.186	.237*
reading to child (03)	.044	.009	.084	.050	.089	-.003	.080	.028	.127
divorce (07)	-.034	-.042	-.021	.015	-.001	.025	-.266***	-.222*	-.314***
N	2562	1330	1232	2582	1349	1233	2564	1339	1225
R^2	.077	.052	.084	.048	.045	.027	.058	.049	.058

Table 2.4: IV estimations for preschool and family daycare

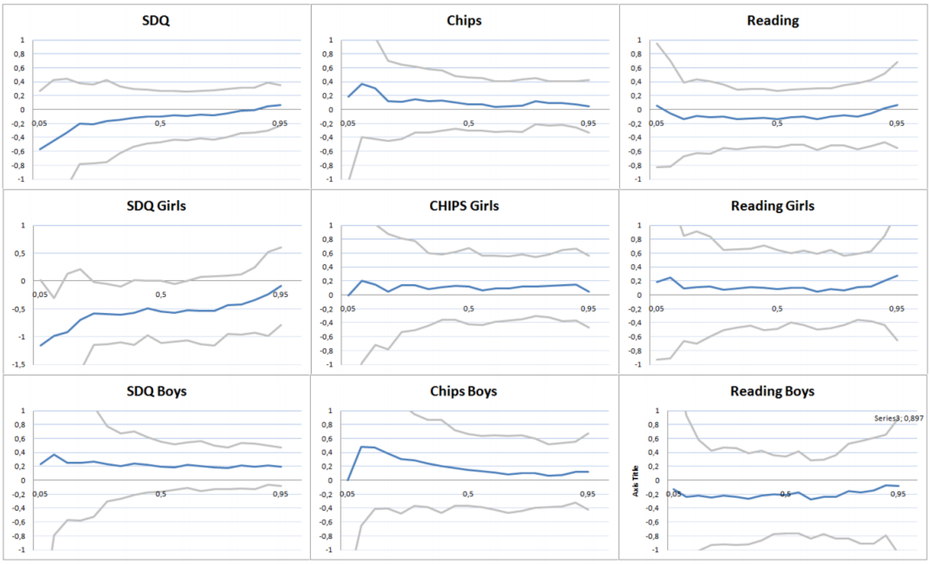

Figure 2.4: non-linear Quantile Regression Coefficients for Preschool on Testscores

close relationship to one caretaker and a relatively small reference group of children. We can only theorize, after looking at our results, that it might be that this type of care arrangement is relatively better for at-problem girls, while at-problem boys tend to benefit more from well-instituted preschool programs. To uncover whether these mechanisms are indeed at play, does however, clearly require further research.

2.5 Conclusion

We find that assessing the effects of universal childcare non-linearly for the case of Denmark, mainly confirms the already established findings of weak to non-existent effects. This, thus stands in contrast to the strong results non-linear analysis yielded for the Norwegian case. It has to be noted however that these results have to be regarded with caution. For one, the sample-size is relatively small. Small sample size becomes an even mopre pressing problem when trying to derive non-linear estimates, which leads to relatively big confidence intervals surrounding the coefficients, in particular at the ends of the distribution. Also it has to be noted that the counterfactual to which we compared high-quality daycare was lesser quality care, which is obviously different than the comparison to no care at all which the Norwegian (Havnes and Mogstad 2010) and Quebecois (Baker et al. 2008) studies looked at. As mentioned another limitation of this study is that due to data-quality problems in the immigrant sample we had to exclude an important group of children from our study, on whom we might have expected preschool programs to have a particularly strong effect. Nevertheless, some important new insights could be won by looking at the non-linear instrumental variable estimates that could neither be seen by looking at mean treatment effects, nor by splitting the sample into subgroups. First we find that much of the strong nonlinear preschool effects we find with standard quantile regression must be attributed to selection, since there is such a strong difference between the non-instrumented and the instrumented quantile regression. We do however also find some support for the hypothesis that effects on cognitive and non-cognitive outcomes should be non-linear, particular in the lowest quantiles. For boys we find the expected nonlinear positive effects at the lower end of the distribution in particular in tests on reasoning ability. However there is also the surprising finding that the non-cognitive outcomes of at-risk girls seem to be more positively affected by family daycare than by preschool, a result we find both in our instrumented and non-instrumented quantile regressions. It might be that the small-group environment of daycare has more beneficial effects on girls than on boys, who might benefit more from professionally trained educators. This however remains a question that demands further research. Further investigation of non-linearity treatment effects in the context of universal childcare is also needed to establish how much external validity the Norwegian results have. In particular a non-linear evaluation of the Baker study of Quebec (Baker et al. 2008) would be useful to see whether the only relative weak existence of non-linearities in the Danish case is due to the circumstances of the study. Studying newer waves of the DALSC data which will come out soon might also lead to more conclusive evidence for the Danish case. Overall looking at distributional effects instead of only mean effects should become an increasingly more important toolkit in the evaluation of large-scale public programs.

3 Rusty Instruments? Revisiting the Twin Approach to Estimating the Relationship between Fertility and Maternal Labor Market Outcomes

Abstract

This paper revisits the link between fertility and subsequent maternal labor-market outcomes. Using panel data for the entire Danish population we are able to present precise estimates, showing the effects of having a child on the following 15 years of the mothers' career. We also show that previous results relying on twin birth as a source of exogenous variation in the number of children are bound to be flawed. We provide evidence that not properly taking into account differing subsequent fertility behavior of twinning and singleton mothers leads to upwardly biased estimates. This bias should be less pronounced when analyzing twinning at higher birth parities. Our estimates show that the effects of children on maternal labor market outcomes are indeed bigger and more lasting at higher parities, cumulatively resulting in an average effect of close to a a year of lost labor due to a birth, over a 15 year time frame. We build a case that the difference in estimates is mainly due to the fact that the instrument loses more sharpness at lower birth parities, even though different marginal effects due to different parities cannot be ruled out. We also present results showing that lower income mothers are more heavily affected in their labor market trajectory.

3.1 Introduction

Over the last century industrialized nations have experienced a vast increase in the supply of female labor, as well as a strong decline in fertility (Smith and Ward 1985). These developments are generally regarded as related. However, pinning down the exact relationship between labor-market outcomes and fertility remains difficult. The decision to have children, as well as most important economic decisions of household members are very likely to be, at least partially, jointly determined. This makes it hard to circumvent the underlying endogeneity problems and to know more clearly how factors, such as labor-market participation and wages, affect fertility and vice versa. While both, the effect of labor-market outcomes on fertility and inversely the effect of fertility on a mothers' labor-market outcomes are of great interest, our work contributes only to the analysis of the latter.

Our empirical strategy relies on exploiting exogenous variation in fertility induced by twin-birth. Using high-quality registry data from Denmark, which provides us with a rich panel data-set of the entire Danish population, we are able to show on a year-by-year basis how fertility shocks develop over time. By fertility shocks we mean the sudden unplanned arrival of another child due to twinning. More precisely we look at the effects of fertility on labor-force participation and gross income in the 15 years after birth. We also look at how the consequences of twinning differ by birth parity, i.e. whether the effects of having an extra child at first birth are similar to those of having an extra child at second birth, and so forth. When doing so, we find that the negative effects of twinning on our outcome variables increase with each of the four birth parities we consider.

There are two possible explanations for this finding. The first is simply that households observed at different birth parities are fundamentally different. This might be the case for a variety of reasons. It could be that the underlying characteristics that lead people to have more children are also in some way related to the strength of the labor-market response to having children. It could also be that children at different parities affect the household economy differently. Maybe each additional child becomes harder to take care off, given that time and financial resources are already constrained by the presence of previous children. On the other hand it could also be that there are economies of scale, making it easier to take care of each additional child. All of these effects add up to an explanation, in which it would be differences between households that explain why we encounter so strongly contrasting effects at different birth parities. The second line of explanation is that the quality of the instrumental variable for properly correcting estimates might differ with birth parity. This implies the reaction to having a child does not differ as strongly with birth parity as our estimates indicate, but instead the estimates for some birth parities suffer from considerable bias. We show that there is a good amount of evidence that is remarkably consistent with the view that the differently sized effects we find are indeed due to the second explanation and that twinning as an instrumental variable works better at higher birth parities. The reason for which we believe the instrumental variable to lose its "sharpness" and to do so particularly in the case of first-birth twins is to be found in what we call subsequent fertility behavior. By this we mean the pattern of births, that come after the particular birth parity for which we compare twinning and singleton mothers. We show that the average difference in children between these two groups of mothers is, unsurprisingly, exactly equal to one child in the year that birth is given but then shrinks to substantially lower values.

When running an OLS regression with a dummy variable for twinning on the explanatory side and outcomes measured several years after birth as dependent variables, the twinning dummy does not capture the effect of one extra child, but rather the effect of whatever is the average numerical difference in children between twinning and singleton mothers at the point in time at which the outcome variable is measured. Thus a coefficient for twinning might capture the effect of almost one extra child on labor market outcomes, in the year after birth was given, but will instead reflect the effect of only halve an extra child several years after birth, making it difficult to compare estimates across time. If the problems that these differing subsequent fertility behaviors cause for our estimates would merely be a question of correctly scaling the estimates, so that the twinning coefficient consistently corresponds to a difference in children that is equal to one, we would be confronted with a rather simple exercise of numerical correction.[1] Unfortunately, however a shrinking difference of children between twinning and non-twinning mothers is invariably linked to the fact that some time after the initial birth, singleton mothers have a different probability of again having young children at home than the twinning mothers we compare them to. Since children tend to have the

[1] As we will see this consequence only affects estimates done via the twins first methodology, Instrumental Variable estimates are accordingly scaled

biggest effect on a mothers' career when they are very young and their mothers actually interrupt working for maternity leave or scale back on the number of hours they work in order to provide more intense maternal care, a differing probability of having young children at home is going to affect the wages and employment we measure at a given point and will thus affect our estimates in a more complex way. This problematic difference in subsequent fertility behavior between the two groups of mothers we compare is much less pronounced at higher birth parities. As we will show, the fertility difference between twinning and singleton mothers remains closer to unity at fourth-birth than it does at third-birth and does so more at third-birth than at second-birth and so forth. It follows that for high birth-parities the distorting effect that subsequent fertility behavior has on our estimates becomes increasingly less problematic. Our estimates show that previous results in the literature on the relationship between fertility and maternal labor market outcomes probably underestimate the depth and the duration of the negative shock for a mothers career that result from having a child. This follows from the fact that in order to estimate long-term effects on maternal outcomes previous studies used either samples consisting only of twinning and non-twinning mothers at firstbirth (Rosenzweig and Wolpin 1980) (Jacobsen et al. 1999) or at secondbirth(Angrist and Evans 1998). A main reason for this focus on low parities was obviously due to the need to obtain a sufficiently large sample of twins. Another weakness of the vast majority of previous estimates is that they are done almost exclusively with cross-sectional data.[2] The authors generally group together individuals from different cohorts in their cross-sectional sample and look at the time that has passed since birth, which does in turn not allow them to properly distinguish between cohort effects and time effects. The panel nature of our data as well as the big sample size allow us to calculate exact year-by-year curves of the effects of fertility changes on maternal outcomes and to control for cohort effects.

In the next section we go on to discuss the previous studies that set out to uncover the relationship between fertility and labor-market outcomes. As a consequence we will also discuss the different attempts to discover sources of exogenous variation in fertility. In the section following after that we then go on to describe our data as well as the situation of mothers in Denmark. The discussion of our main hypotheses and of the different

[2]A notable exception is the use of panel data by Carrasco (Carrasco 1998)

empirical strategies we employ to look at the effects of twin-birth on fertility will make up the next section. We end this with a discussion of the importance of accounting for birth-order in our estimates. Then the main results are presented, followed by additional results and a series of checks on their robustness. Finally, we conclude by placing our findings into the context of the previous literature.

3.2 Literature

Ideas of how fertility affects household wealth accumulation go back to Malthus. His basic equilibrium model postulates that as households grow richer, they will keep on having children until every economic surplus at their disposition is evaporated and the household moves back to living at subsistence level.
Shortly after Malthus postulated his thesis however the Malthusian model ceased to be an adequate description of the industrialized world. Instead living standards per capita kept rising and fertility did not keep up.
In particular within the last century the role of women in most advanced economies underwent rapid changes again, with an increasing labor-market integration of women being combined with often decreasing fertility rates. This led to a renewed interest in the interaction between fertility and the economic outcomes of households, or more specifically the outcomes of mothers. The big picture is further complicated by the fact that among industrialized nations the previously negative relationship between female labor-market integration and overall societal fertility is being reversed. Among OECD countries it is the likes of Sweden, Denmark and the US, which have been at the forefront of female labor market participation, that are suddenly displaying much higher fertility rates than more traditionalist societies, like Italy or Spain (Feyrer et al. 2008). While these broad macro-trends may show a reversal in the relationship between a countries' female labor-force participation and fertility, at the micro-level, the historical, theoretical and empirical evidence still suggests that if you look at the labor-market trajectory of an individual mother, there is a negative relationship between fertility and a mother's labor-force participation.

This crude association is supported by a series of detailed historical studies. For example Goldin (Goldin 1995) has shown for 5 cohorts of female American college graduates between 1910 and 1991 that combining the founding of a family with children and a career has consistently proven difficult for mothers.

Theoretically it is not entirely clear in which direction we would expect a fertility shock, meaning an unexpected positive increase in the number of children, to affect a mothers labor-market outcomes, since the income effect (children are expensive) might push her to work more, while the substitution effect (children do consume time, thus raising the reservation wage) should be in the opposite direction. But most of the theoretical literature tends to stress predominantly the effect that after the birth of children, the value that mothers assign to non-work time rises. For example, Gronau (Gronau 1974) argues that the dominant labor-market effect of children is their effect on the price of time.

Finally, there is a vast amount of empirical studies showing a negative relationship between child-birth and the mother's labor-market participation and income. Waldfogel provides a thorough survey of this literature (Waldfogel 1998). However, as Browning noted in his review on children and the economic behavior of households in 1992, few of the studies done prior to his survey dealt with the endogeneity problems complicating the relationship between fertility and maternal labor-market outcomes in a satisfactory way. Thus one had to be cautious about drawing inferences (Browning 1992).
Agnrist captured the problem well with his statement that nothing illustrates the inherent endogeneity problem more clearly than the fact that economists run regressions with labor-market outcomes as the dependent variable and fertility variables on the independent side, while demographers turn the equation around and explain fertility outcomes, by using labor-market characteristics (Angrist and Evans 1998).[3]
Since Browning's critique, there has been an increasing number of studies aiming to look at the relationship between fertility and labor market outcomes in setups that allow for causal inference.

The attempt to achieve valid causal inference has mainly been pursued via instrumental variable estimation techniques. However the search for variables that correlate with fertility but have no effect on labor-market outcomes, which is the requirement for a valid instrument in this case,

[3] A notable exception was Mincer who insisted on not including fertility variables in labor market outcome regressions (Mincer 1963)

is a complicated and often elusive quest. Among the earliest suggested instruments for fertility were the mothers ideal family size as expressed in a survey and the mothers religious affiliation (Cramer 1980) as well as the country of origin (Schultz 1977). Rosenzweig and Wolpin were the first to use the "natural natural experiment of twin birth", which up to date has remained the most prominent instrument to estimate the causal effect of fertility on labor supply (Rosenzweig and Wolpin 1980). Jeff Groger and Stephen G. Bronars then went on to use the occurence of twins as a source of exogenous variation in Welfare Payments for mothers (Grogger and Bronars 2001). Jacobsen, Pearce III and Rosenbloom re-estimated the effects of twin-births on mothers' labour market outcomes in more detail using large US census samples (Jacobsen et al. 1999). Angrist and Evans (Angrist and Evans 1998) used twin-births to look at maternal and paternal labour market outcomes, but also introduced a new instrumental variable. Their approach uses a couple's preference for having children of mixed gender. The fact that in most industrialized nations families, whose first two children have a mixed gender composition are less likely to have a third child than families which have two boys or two girls provided them with an additional source of exogenous variation in a couple's number of children. Since these estimations calculate the effects of a marginal extra third child Angrist and Evans go on to check the results obtained by the new instrument by comparing them to twinning at second birth. Carrasco estimated effects on labour force participation using sex-composition in a panel setup. Aguero and Marks introduced yet a new instrument, trying to identify infertile women in health surveys and thus providing new estimates for Labor Force Participation using those as an instrument (Aguero and Marks 2008). Finally Simonsen and Calceres used variation in the number of children via twinning to look at an entire array of maternal health and wellbeing outcomes (Cáceres-Delpiano and Simonsen 2012).

These studies have overwhelmingly found that fertility shocks have a negative effect on mothers labor market outcomes, but that the effect is much smaller than a standard OLS estimation would imply.[4] Also, effects have been found to be non-persistent. The estimates on when the effects of an extra child perish, differ, but they tend to generally range between 2 and 13 years.

A second line of research in the twinning literature has not focused on

[4] An exception is the Aguero and Marks study which finds no effects (Aguero and Marks 2008)

maternal outcomes but instead looked at the effects that the unexpected arrival of an additional sibling has on child outcomes. These studies have predominantly tested the quantity-quality model of children going back to Gary Becker and Lewis (Becker and Lewis 1973) as well as to Becker and Tomes (Becker and Tomes 1976). This literature inclueds the work by Black, Devreux and Salvanes looking at effects of additional siblings on children's educational attainments in Norway (Black et al. 2005), Caceres looking at school outcomes (Caceres-Delpiano 2006) and Angrist, Levy and Schlosser Testing a series of human capital related outcomes such as earnings and education in Israel. (Angrist et al. 2010a)

Most of the more recently written papers using twinning as a source of exogenous variation, look at how the effect of an extra child varies at the margin, meaning they analyze the effect of an extra child via twinning for a given birth parity. Good examples include the work by Black, Devreux and Salvanes (Black et al. 2005), who restrict their sample to families having given at least n births and who do calculate seperate parameters by birth parity. Another example that does this is the work by Angrist, Levy and Schlosser (Angrist et al. 2010a). It has to be noted that all of these papers do look at child outcomes however. In the literature on maternal outcomes, we are only aware of the work of Simonsen and Calceres (Cáceres-Delpiano and Simonsen 2012) on maternal health and well-being as actually analyzing the effects at the margins for different birth parities. The entire previous literature on maternal labor-market outcomes normally just looks at twinning at first birth (Rosenzweig and Wolpin 1980) (Jacobsen et al. 1999) (Grogger and Bronars 2001) or at second birth (Angrist and Evans 1998).[5] Looking at twinning by birth-order means acknowledging that the effects of a child might vary according to the margin, which in turn implies that what our twinning studies allow us to say about what might be the most important transition, namely the one from having no children to having one is very limited. As Waldfogel notes in her survey it is at this transition that we actually observe the biggest wage differences (Waldfogel 1998) and it is also the transition were selection effects might be strongest, thus one has to be aware that twin-studies are of somewhat limited use to assess the transformation from being childless to having one child.

Many of the more detailed questions about the motherhood wage penalty

[5] Angrist looks at second birth twinning to ensure comparability to the sex-composition instrument he introduced int the same paper (Angrist and Evans 1998)

have thus far been almost exclusively explored in non-experimental studies. These studies obviously suffer from the fact that they have to try to control for all relevant factors that account for the differences between mothers having children and those that do not. Still on many issues the findings of this literature are more differentiated than those of the twinning literature. For example Budig and Hodges have shown that mothers at the bottom of the income distribution tend to pay a larger penalty for motherhood than those at the top (Budig and Hodges 2010), while others argue that highly skilled mothers tend to pay a bigger penalty (Wilde et al. 2010). Of particular interest for the issues we explore in this paper is a recent study by Kahn, Garcia-Manglano and Bianchi (Kahn et al. 2014), which stresses many points that are similar to the ones we attempt to make. The Kahn et al study also looks at long-term labor market trajectories of mothers after birth and just like we do, differentiates the effects according to birth parity. What they find is that mothers' careers tend to be severely impacted shortly after child birth but tend to normalize at higher ages. They also find that there is a significantly bigger wage penalty at higher birth parities. This is important to bear in mind. Our paper argues that much of the difference in the motherhood wage penalty is due to the fact that at lower parities the instrumental variable estimator suffers from bias. However the fact that a non-experimental study came up with similar findings regarding parities offers another possible interpretation, which is that mothers with more children might simply become significantly less attached to the labor market and are likely to truly suffer a bigger wage penalty than those at lower parities. We will discuss how to strike a balance between these two interpretations in our conclusion.

While we have thus far mainly stressed the advantage of twinning instrumental variable studies over ones that seek to estimate wage penalties using the standard approach of trying to control for as many maternal characteristics as possible, one also has to be aware of some of the potential weaknesses of the twinning approach. In particular the question of external validity is important in this respect. To put it simply, the question arises whether having twins is really that similar to having two singleton children. Black, Devereux and Salvanes discuss this in detail in their work on the birth weight of children (Black et al. 2007) and argue that there are significant differences between twins and singletons with respect to birth complications, gestation, health and one-year mortality. This already indicates that twin-birth is in many ways a very different event than the birth of singletons. For our work the question arises whether having two children at once might place a whole different set of demands on the mother than would be the case if she had given birth to two singletons instead. The time budget of the mother might be much more severely constrained in the case of twins. On the other hand it might also be that there are important economies of scale associated having two equally aged children at once, allowing the mother to perform many of the care activities for both children at once. Much more detailed studies looking at data like time-use and consumption patterns of twin-mothers would be needed to cast some light on these questions and to give us a better sense of how much external validity twin studies actually contain. When interpreting the results it is important to bear in mind that external validity of estimates on twins cannot simply be taken for granted.

Our paper falls firmly into the camp of the literature looking at the effects that variation in the number children has on parental outcomes, not the one looking at changes in child outcomes. Having panel-data we can do so more precisely than previous studies and show exactly how the effect of an extra-birth evolves over time. While this has become common in the recent literature on child-outcomes, we do, to our knowledge, present the first study taking a detailed look at how shocks on fertility differ in their effects on labour-market outcomes differ by birth-order. Further we expand on the literature by providing evidence that the standard twinning instrument suffers from serious problems when used to look at long-term outcomes. We also present empirical strategies aimed to address this problem.

3.3 Data and Background

We use high-quality Danish Registry data to estimate the effects of fertility on mothers' labour supply and gross income. This allows us to derive estimates for a sample comprising the entire Danish population. Apart from granting us a very big samplesize there are several advantages to the data. For the cohorts of mothers we look at, which are all women that gave birth between 1980 and 1992, Denmark still had a relatively high degree of ethnic homogeneity. The immigrant share of the Danish population was under 3% (under 4% including second-generation immigrants) in 1980, which is the time at which we start looking at our first cohort and under 4 % (under 5% including second-generation immigrants) in 1992 the year

Birthtype	Frequency	Percentage
Singletons	1,845,945	99.02
Twins	18,106	.97
Triplet	214	.01
Quadruplets	11	.00
Quintuplets	3	.00
Sixtuplets	2	.00
Total	1,864,281	100.00

Table 3.1: All Danish Births 1980-92

in which we start following our last cohort of mothers (Liebig 2007). Add to that, that over half of the immigrants living in Denmark at the time were from other OECD countries (Liebig 2007). Since, as we wil discuss, the probability of twinning can vary with ethnicity, this homogeneity of the population assures us that our results are not going to be substantially affected by the bias this might introduce.

Further, fertility in Denmark has been far more constant than in many other OECD countries. In contrast to most other OECD countries it actually experienced a small rise in fertility in the period from 1980 to 1992 and remained constant after that (D'Addio and d'Ercole 2005), as can be seen in the small differences in total realized fertility between our 1980 and 1992 birthcohorts (see Tab. 3.2). This means that when pooling data from different birthyears, abrupt changes in fertility patterns and associated changes in the selection into and out of fertility are not going to significantly affect our estimates.
It has to be noted that there has been an ongoing change towards higher educated women becoming relatively more fertile in Denmark though and in an analysis of several European countries done by Esping-Andersen it was the only one were having a child actually predicted a lower probability for a given family to fall below the poverty line (Esping-Andersen Esping-Andersen)

It also has to be noted that Denmark has extremely generous legislation for assisting mothers (OECD 2012). There are 18 weeks of paid maternity leave. Further Denmark actively assists mothers who gave multiple births.

In 2010 DKK 8,024 a month was paid for each additional child, which was due to multiple births. The grant is paid quarterly until the children are age 7. Also, a lump sum of DKK. 46,214 is paid at birth. This means that the twinning mothers receive a significantly greater amount of financial help relative to the singleton mothers. It is worth noting though that the additional support, that twinning mothers receive relative to singleton mothers does not vary with birth parity. This is important since a fair amount of our conclusions depend on analyzing the differing behavioral response of mothers to an additional birth according to birth parity. If the financial situation of the twinning mothers relative to singleton mothers would strongly differ with birth parity this might in part be the driver of our results. While a case might be made that the generous payments mothers receive might work as a disincentive to start working again, it has to be noted that the high-quality early childcare in Denmark is much more generous than it is for example in the United States (Esping-Andersen et al. 2012), which probably makes it easier for mothers to combine having children with work than it is in most other countries. It is also worth noting in this context that Denmark has consistently had one of the highest female employment rates in the OECD (Jaumotte 2003b).

The Danish Registry records key demographic and economic variables for the entire population on a yearly basis. In the version of registry data we had access to, persons were recorded starting at age 15. The number of children that a person has are recorded according to their age in the registry. This does not allow for exact twin identification, since adoptions or giving birth twice a year are not identifiable in the data. Thus, to correctly identify twins we only considered individuals which actually appeared in the registry themselves. We could then use information on the exact birth date as well as on the identity of mother and father to identify twins. We merged these identified twins to the data of the mothers in the year they gave birth. This means that we could only identify twin births retrospectively, once the twins actually reached age 15 and were thus recorded as individuals in the registry data. Since we had the registry data available up to 2007 we were able to identify twin births up to 1992. Table 3.2 reports the descriptive statistics for our sample. The variable on the total number of kids reports the completed fertility as measured in 2007. It can also be seen from that table that the average employment and probably as a consequence, the average income of 1st birth mothers is higher in the year before birth than that of second and third birth mothers. In both cases there is thus actually more potential for a drop in employment and income, in our firstbirth sample than in the second or third birth sample. It is also noteworthy that the relative number of twins among total births seems to have gone up significantly in the 1992 data. This could on the one hand be due to the fact that people are giving birth at a higher age, which tends to increase the probability of twinning. It could also be due to the fact that in the later years of our sample in-vitro fertilization actually starts being used more widely. For this reason we included a version of our estimates using only the earlier half (1980-1986) of the cohorts we follow to reestimate our results in the section on robustness checks.

Since we are interested in how the effects of fertility play out after birth, we recode time, which is measured in years in the case of the registry data, as $t \in [0, 15]$ to capture the time that has passed since birth was given. Thus for example $t = 1$ means that a year has passed since the birth we are analyzing for a given mother. In our estimations we pool all the years we have available, thus a mother that gave birth in 1981 and one that gave birth in 1989 would both appear in our estimation estimating effects two years after birth at $t = 2$ with their recorded data for 1983 and 1991 respectively. Therefore we included a set of dummy variables for the different birth years in out models to control for time effects.

	1st birth		2nd birth		3rd birth		4th birth	
	twins	*singletons*	*twins*	*singletons*	*twins*	*singletons*	*twins*	*singletons*
Total Kids								
1980-92	2.52	2.20	3.20	2.46	4.12	3.31	5.22	4.32
1980	2.50	2.12	3.17	2.38	4.18	3.25	5.00	4.3
1992	2.46	2.20	3.22	2.48	4.11	3.34	5.23	4.33
Mothers								
Age at Birth								
1980-92	27.23	25.81	29.36	28.68	31.88	31.44	33.20	33.20
1980	25.82	24.78	28.31	27.87	31.11	30.73	32.60	32.31
1992	28.64	26.81	29.77	29.41	31.97	31.80	33.53	33.18
Educ in Ys								
1980-92	12.30	11.99	12.13	12.00	11.50	11.62	10.75	10.93
1980	11.84	11.52	11.56	11.58	11.06	11.15	10.1	10.2
1992	12.59	12.36	12.35	12.35	11.69	12.01	11.07	11.19
Inc. Y bef. Birth								
1980-91	125,676	112,889	119,866	114,689	106,502	105,147	86,563	90,141
1980	80,821	72,670	71,776	68,899	56,728	58,464	43,396	48,541
1991	154,977	141,363	151,546	147,251	145,349	138,240	95,803	115,928
Emp Y bef. Birth								
1980-91	.803	.770	.776	.751	.689	.684	.532	.549
1980	.789	.780	.736	.750	.676	.649	.500	.530
1991	.814	.745	.731	.722	.705	.672	.48	.523

Fathers								
Inc Y bef. Birth								
1980-91	173,936	157,753	188,193	181,713	194,881	197,532	186,874	200,357
1980	117,290	106,184	130,730	126,189	138,209	136,961	117,441	133,304
1991	217,763	192,905	231,961	222,196	231,592	236,282	205,190	252,846
Emp Y bef. Birth								
1980-91	.866	.863	.909	.898	.886	.888	.826	.837
1980	.862	.859	.919	.902	.955	.902	.846	.857
1991	.865	.832	.904	.871	.839	.857	.760	.809
N								
1980-92	3,287	337,215	3,004	271,503	1,051	91,199	235	20,381
1980	197	25,783	228	21,161	84	6,915	10	1,429
1992	421	29,338	303	24,758	95	8,807	26	2,031

Table 3.2: Mean of Mother Characteristics Recorded at Year of Birth

3.4 Twinbirth as a Natural Experiment: Methodology and Identification

Twinbirth is generally treated as a classical natural experiment. The motivating idea is that by a stroke of chance a mother gives birth to two kids instead of one. The truth is that twinning is the result of a biophysical process that is unfortunately not entirely random. Thus, we know of at least three caveats that should be kept in mind when looking at the effects of twinbirth.

First, while the probability of getting monozygotic twins is relatively stable across age, the medical literature tells us that the probability of byzigotic twin-birth strongly increases for older women (Hoekstra et al. 2008) (SMITS and MONDEN 2011). This is generally accounted for by including controls for age and age squared in the estimation models. Second, as mentioned before the probability of twinning is influenced by ethnicity. Third, in-vitro fertilization (IVF) can affect twinning in several ways. In the early stages of the technology it simply led to an increased probability of twinning relative to natural fertilization. More recently it has increasingly become possible to offer mothers the choice to have twins or not. This means that with the availability of in-vitro fertilization, we either have an endogeneity problem due to increased twinning among mothers who chose to undergo that procedure, or an even worse endogeneity problem because twinning to some extent becomes a choice-variable. The first in-vitro fertilization ever took place in 1978 and in the first halve of the 1980s employment of the technique was still extremely rare in Denmark (Westergaard et al. 1997), however since we also follow maternal cohorts giving birth in the early 1990s, worries about IVF affecting our results might be valid. We address these concerns in the section on robustness checks by looking at whether our main results hold for a reduced sample including only mothers that gave birth no later than 1986. As newer data becomes available the potential pitfalls of IVF for research using twinning as an exogenous source of fertility variation are bound to become more pronounced.

The role of factors altering the probability of twinning has been thoroughly explored by the medical literature (Hardin et al. 2009) and is generally acknowledged among economists and demographers. What has received far less attention is how the adaption in fertility behavior that follows twinning is bound to affect estimates relying on twinning as a source of exogenous variation. Most works acknowledge the fact that mothers giving birth to twins tend to adapt their subsequent fertility behavior (see for example Angrist et al. (2010b)). More precisely it is generally pointed out in the descriptive statistics on total amount of children that twinning and non-twinning mothers have a difference in completed fertility rates that is much smaller than 1. This is because a mother that gives birth to twins at the *n*th birth is less likely to have further births after that than a mother that gives birth to a singleton at her *n*th birth. A thorough discussion of how exactly this difference in subsequent fertility behavior is bound to affect estimates is missing from the literature however.

Another factor to be aware of is that the occurrence of twinning is a change in number of children that happens at the margins. So twinning of a mother at firstbirth might not have the same effect as twinning of a mother that already had two kids. Lumping together twinbirths that occurred at different parities is not necessarily problematic in the sense that it should still give you a weighted average of the effects of an extra-child. If marginal effects do however significantly differ from each other, important information would be lost.

3.4.1 Hypotheses

From our discussion of how subsequent fertility behavior affects twinning we derive the three following key hypothesis

1. We expect that fertility after the *n*th birth of twinning mothers is substantially lower than that of non-twinning mothers for all *n*.
2. This difference in subsequent fertility behavior should be less pronounced at higher birth parities or for mothers giving birth later in their lifes
3. The differences in subsequent fertility behavior are bound to upward bias our estimates, simply because when looking at labor market outcomes later in life non-twinning mothers are more likely to have children at home, which leads to lower incomes and employment in our comparison group. We thus expect to find stronger negative effects of twinning at higher birth parities or for births that happen later in life.

3.4.2 The effects of twinning on subsequent fertility

The interest of this paper and of the previous twinning literature that we expand on is to derive estimates about the effects of fertility on long-term

labor market outcomes. However, as we have pointed out, an important and insufficiently discussed consequence of twinning is that it does not only affect a womens labor market development by changing the number of children she has, it will also affect her subsequent fertility decisions. Think of a mother who had planned to give birth to two children, but gives birth to twins at her first birth. She might simply stop giving birth after that event, whereas she might have given birth for another time had she given birth to a singleton.
To treat this more formally let us denote the number of children mother i has at after t years have passed since her nth birth as C_{itn}, let us also denote giving births to twins as T_{in}. The idea of using twinning as a source of exogenous variation in fertility is that you assume that a mother who gave birth to a twin has a child she would otherwise not have. So ideally for twinning to be a perfect treatment we would want

$$(C_{itn}|T_{in}=1)-(C_{itn}|T_{in}=0)=1\ \forall t \tag{3.1}$$

This means that the number of children of a woman who gave birth to twins relative to the same woman who gave normal birth, has truly been raised by one and stays at 1, for all the time t that passes after birth. However, as our simple thought experiment shows this is not what we expect to happen. Mothers that have reached the number of children they wanted to have via twin-birth, without having surpassed it, might restrain subsequent fertility, relative to the case where they had no twins. Thus, with the passing of time the fertility difference between twinning and non-twinning mothers is bound to fall below 1. More precisely we postulate the following assumptions about the relationship of twinning and the total number of children a mother goes on to have:

1. $(C_{itn}|T_{in}=1)-(C_{itn}|T_{in}=0)=1\ \forall n>0$ at $t=0$
2. $0<(C_{itn}|T_{in}=1)-(C_{itn}|T_{in}=0)\leq 1\ \forall n>0$ at $t>0$
3. $\partial(C_{itn}|T_{in}=1)-(C_{itn}|T_{in}=0)=1/\partial t\leq 0\ \forall n>0,t>0$
4. $\partial(C_{itn}|T_{in}=1)-(C_{itn}|T_{in}=0)=1/\partial n\geq 0\ \forall n>0,t>0$

We know from surveys that many women do not actually achieve their desired fertility. In Denmark the ratio of actual fertility to desired fertility is relatively high at 0.8, compared to an EU average of 0.6 (Esping-Andersen Esping-Andersen). In order to illustrate the logic of our four assumptions

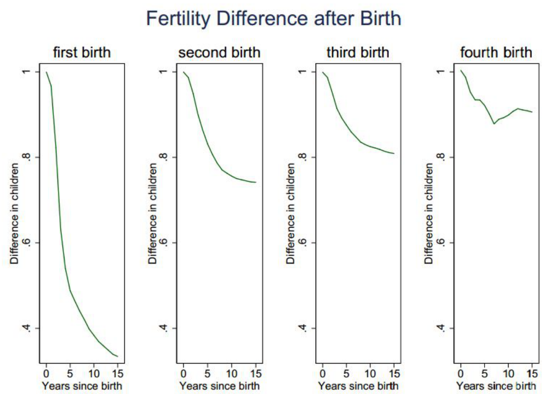

Figure 3.1: Difference in Fertility Between Twin and Non-Twin Mothers

the idea of desired fertility proves to be a valuable concept.

The first point (1) we make is the trivial statement that, at the time of giving birth, the difference in children between mothers that had twins and the ones that did not is exactly one (for simplification we abstract from triplings, quadruplets, etc.).

The second point states that this difference in realized fertility might fall below one but not rise above it as time passes. It also should not fall below 0. This is because mothers, that gave birth to twins, might have given birth to a child that they planned to have in the future. In the extreme case that all of the twinning mothers had a child, which they would have also given birth to had they not twinned, the difference in children between twinning and non-twinning mothers would fall to 0 for a sufficiently large t. In the other extreme case that none of the mothers were actually planning to have more children, than the one they just gave birth to, the difference would remain stable at 1 as time passes. The reality is bound to lie somewhere in between.

The third point follows from the same logic. Mothers that were planning to have another birth after their nth birth, might have been planning to do so at different times t. So as time passes, the singleton mothers that were planning to have $n+1$ children will at different points in time t go on to have another child, while the twinning mothers desiring the same number of children will not have to give birth again. Thus, the difference in the number of children between twinning and singleton mothers is bound to be monotonically decreasing in t.

The most complex point is the fourth one. It states that the difference in fertility between twinning mothers and their singleton counterparts is going to be bigger, the higher up we move in the birthorder. In other words the shrinking of the initial difference of 1 that happens over time (point 3.), is going to be most pronounced when looking at the effect of twinning at first birth, less pronounced at second birth and so forth. Thinking about desired fertility again illustrates the point: A mother having twins at first birth reduces subsequent fertility, in any of the cases where she wanted to have two children or more, a mother having twins at second birth would only be induced to reduce fertility in the cases where she aimed to have three children or more, etc. Thus the assumption behind the fourth statement is that the ratio of mothers who want $n+1$ children among those who already have n children is decreasing in n . This is an assumption that fits observed fertility data in most industrialized countries. The ratio of women having $n+1$ children to women having n children, is generally decreasing in n for $n > 1$. For our purposes this implies that, even though we expect the fertility difference between twinning-mothers and non-twinning mothers to decline over time, this decline should be least pronounced for twinning at higher birth-orders.

In table 3.2 you can see our best approximation of completed fertility for the mothers in our sample. In our case this is the the total number of children a mother had up to 2007, which is the last year we have data for. This restriction might lead to the completed fertility for later cohorts to be slightly underestimated, given that a woman giving birth in 1980 had her completed fertility measured 27 years after that birth while a mother giving birth in 1992 only 15 years after that birth. However the relatively low fertility of women over age 35 in our sample assures us that the effect that the truncated data has on our estimates of final fertility should not be too big. In the table you can see the average completed fertility for twinning and singleton mothers at different birth-parities. We included the mean for the full sample, which includes anyone who has given their first-, second, third, or fourthbirth in the years between 1980 and 1992. We also provide summary statistics for the cohorts giving birth in the first (1980) and last year (1992) of birth for which we followed mothers in our sample, to provide a sense of change over time. We can see, that the difference in completed fertility between twinning and singleton mothers is indeed between 0 and 1. We also see that the difference does, as we expected, stay larger for births at higher parities. The difference in completed fertility between twinning and singleton mothers is around .3 if the twinning

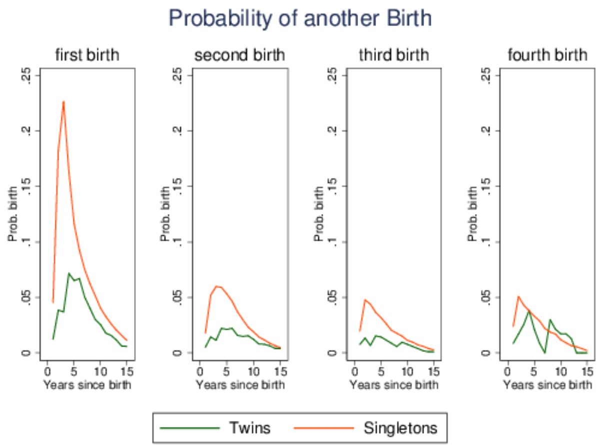

Figure 3.2: Probability of giving birth twin and nontwin mothers

happened at firstbirth, but instead between .7 for second births, .8 for third births and .9 at fourth births.

Fig. 3.1 offers a more detailed view of how the difference between twinning mothers and their non-twinning counterparts evolves over time for the 15 years after birth. The estimations are done for all mothers having given their *n*th birth between 1980 and 1992. Again, we can see all four of our predictions on subsequent fertility behavior confirmed in these graphs. The difference between twinning and non-twinning mothers always starts at 1 and then falls to values between 0 and 1. Also, with the exemption of fourth birth, for which, due to the small sample size estimates are much noisier, the differences between twin- and singleton-mothers are monotonically decreasing in time. Most importantly, the decline in fertility differences is much more pronounced among firstbirth mothers than among those giving birth at higher parities. The higher up we move in birth parities the closer the fertility difference remains to 1. This graph has important implications for how to interpret the outcomes of models, using twinbirth as a source of exogenous variation in fertility. The first implication is that when estimating the effects of fertility on long-term outcomes, via the fertility variation induced by twin-birth, we have to be aware that simple OLS estimates using twin-birth as an exogenous source of variation do not capture the effect of one extra child, but rather, depending on birth-order and on the time that has passed since birth, the effect of a fraction of an extra child that can be as low as .4 extra children in the case of the firstbirth sample . This means we have to be careful when comparing coefficients obtained by these OLS regressions for outcome variables measured at different points in time. A bigger coefficient for the short term effects of twinning might in this case either mean that one extra child really has a bigger effect on a mother in the short term or it might be capturing the effect of a bigger difference in the number of children between twinning and singleton mothers in the short term than in the long-term.

If we could however expect that a reduction in the fertility difference between twinning and singleton mothers would simply lead to a proportional reduction in the coefficients we estimate for twinning, the solution to our problem would be a relatively easy numerical adjustment, namely dividing the coefficient by the fertility difference. As we will discuss, instrumental variable methods do indeed correspond to OLS coefficients that were adjusted for fertility differences in the case of twinning. Unfortunately thinking about the implications of these differences in fertility behavior on the life paths of mothers in the twinning-group and those in the counterfactual singleton group makes it clear that another much harder to account for, problem follows from the different subsequent fertility behaviors. They do not only imply a shrinking of the difference in children during the years after birth. More importantly and as a direct consequence, they imply very different probabilities of having small children at home at given points in time. In general the need children have for direct, intensive maternal care is much higher in the early years of life. It is thus during the time where children are very young that maternity is most disruptive to a woman's career and where there is the highest likelihood that mothers might make decisions such as taking time off from work, switching to less demanding careers or reducing the amount of hours worked.

Figure 3.2 graphs the probability for twinning (blue) and singleton (red) mothers to give birth again, for every year after their *n*th birth. The differences in "subsequent fertility behavior" shown in those graphs are in a way the other side of the coin of the reduction in the difference of children shown in Figure 3.1. The graph illustrates very clearly that after the children for which we compare twinning and singleton mothers have grown out of the most disruptive phase of very young childhood, the singleton mothers have a much higher probability of having very young children at home again due to their higher propensity to give birth again.

This higher probability of having young children at home is going to depress the average wages and employment of singleton mothers relative to those of twinning mothers. Remember that our estimates of the effects that an extra child has on maternal wages and employment after t years are entirely based on comparing the recorded differences between twinning and singleton mothers in time t. If at this point in time singleton mothers are on average more likely to have very young children at home this is going to depress the relative average wage we measure for them substantially. Thus the wage difference we find between singleton and twinning mothers is not only a function of the exogenously induced difference in fertility due to twinning but also of the subsequent difference in fertility behavior.
This is bound to upward bias our estimates of how twinning affects maternal long-term labor market outcomes. Our standard twinning regression is set up in a way that it assumes differences in earnings that arise between twinning and singleton mothers (after controlling for age) to be due to the fact that twinning mothers had an extra child at $t = 0$. If differences in earnings are also driven by differing fertility behavior after $t = 0$ our estimations will falsely attribute those differences to be the direct effect of an extra child via twinning as well. Since normally we expect twinning mothers to have lower earnings and employment due to their fertility shock, the consequence of "subsequent fertility behavior", which will in turn depress the wages of singleton mothers relative to the twinning ones, is going to be that the negative consequences of a fertility shock will appear less profound than they actually are. Our estimates will turn out upward biased. The more time passes, the harder it will be to disentangle to what extent the differences in wages we observe are driven by the exogenous variation in children due to twinning and to what extent they are driven by the consequences of subsequent fertility behavior. Thus with time our instrument will have a tendency to become increasingly "rusty" in properly identifying the wage and employment effect that is due to our initial exogenous change in fertility.

Trying to solve the problem of subsequent fertility behavior, by controlling for it in our regressions is not a feasible solution when attempting to derive causal inference. For one, we would need to have good structural assumptions as to how exactly a child affects a mother at different ages to do so. Further, there is going to be selection of mothers into having another child, thus trying to devise a solution to this problem via control variables, would lead to a renewed confrontation with all the endogeneity problems surrounding fertility that we tried to circumvent initially by looking at twinning. However Figure 3.2 shows that, as we formulated in our second hypothesis, the difference in subsequent fertility behavior between the twinning mothers and non-twinning mothers becomes much lower at higher birthorders. Thus at higher birth-parities our instruments are prone to become less "rusty" and our long-run estimates should be less upward biased than they are at lower parities. Should we find substantially bigger effects of additional children on maternal labor-market outcomes, this would be a very good indicator that our worries about downward bias might be warranted.

3.5 Twins First Methodology

There are different ways of modeling the effects of twinning on subsequent maternal outcomes. The twins-first approach was originally employed by Rosenzweig and Wolpin and has been in use ever since (Rosenzweig and Wolpin 1980). The argument they made is that simply comparing mothers that gave birth to twins at any parity, to mothers that did not get twins would introduce selection problems into the estimates. This is because mothers who get more children will have a higher chance of twinning at some point. To avoid this selection the authors restrain themselves to compare twinning mothers to mothers of singletons only at first birth.
Twinning provides an exogenous source of variation in fertility and the regression setup is thus to include a twinning variable in an OLS regression. Obviously the restriction of the sample to first-birth mothers, that was traditionally done to avoid selection issues related to the probability of further births can be generalized to other birth parities. One can just as easily imagine running a twins-first regression on a sample restricted only to second birth or third birth mothers and this is indeed what we will do. We define the labor market outcomes of mother i measured t years after her nth birth as Y_{itn}. We thus estimate the twins first approach as follows

$$Y_{itn} = \beta_{0tn} + \beta_{1tn} Age_{itn} + \beta_{2tn} Age^2_{itn} + \beta_{3tn} Twins_{ni} + \varepsilon_{tni} \qquad (3.2)$$

The coefficient of interest in this case is β_{3tn} which captures the effect of twinbirth at parity n, with t years having passed since the nth birth was given. When looking at how the effects of twinning at different birth parities develop over time we estimate this equation separately for each parity and time-period. We thus obtain a set of $n \times t$ coefficients.

When interpreting the coefficients we obtain for twinning in this setup, we have to take into account that tthey can only be interpreted as the effect of giving birth to twins, which is not at all the same, as the effect of having an extra child.
As made clear by our discussion of post-birth fertility behavior, the further on we move in time since the birth occurred, the smaller the difference in children between the twin mothers and the singleton mothers becomes. At first-birth the twins-first coefficient would capture a difference around 0.9 children at $t = 1$ and below 0.5 children at $t = 5$. In the twins-first methodology there is neither an adjustment for the twinning coefficients to accurately reflect the fertility difference between twinning and non-twinning mothers, nor is it able to account for subsequent fertility behavior. Indeed in some of the previous papers estimating the long-term effects of twinning on maternal labor market outcomes, the authors do at times find positive effects of twinning on long-term labor market outcomes of mothers (Jacobsen et al. 1999). While this is not theoretically impossible, it is still rather unexpected. Taking into account the fact that the strongly differing subsequent fertility behavior in a sample of only first-birth mothers would lead to substantial upward bias offers an alternative explanation for these findings that might be more plausible from a theoretical standpoint.

3.6 Instrumental Variable Estimation

Another technique for estimating the effects of fertility on different types of outcomes is the IV approach. The occurrence of twinning at the nth birth is in this case viewed as an instrument for the total number of children a mother has. Our second step estimation thus looks at the effect that the number of children a mother has at time t has on her labor-market outcomes. And in the first stage we instrument the number of children a mother had with a variable capturing whether she twinned at nth birth. In order to estimate the marginal effects of an additional child at each parity we adopt an approach to restrict our sample that is similar to the one we use when estimating via twins-first. When estimating the effects of an extra child on labor market outcomes t years after birth we thus restrict the sample to all the mothers who had their nth birth at $t = 0$. Then we instrument the recorded number of children in time t with whether the mother twinned in $t = 0$. We estimate these regressions separately for all fifteen years t we consider after birth was given and for the four first-birth parities n. The generalized version of the first stage regression of the IV estimations that we run thus looks as follows[6]:

$$\hat{C}_{itn} = \gamma_{0tn} + \gamma_{1tn}Age_{itn} + \gamma_{2tn}Age_{itn}^2 + \gamma_{3tn}Twins_{ni} + \mu_{tni} \qquad (3.3)$$

The second stage regression then looks similar to 3.2, but instead of the $\beta_{3tn}Twins_{ni}$ we now include $\hat{C}_{itn}$. Note that the coefficient γ_{3tn} captures the effect twinning at the nth parity has on the number of children a person has at time t conditional on age. Angrist (Angrist 1991) has shown that in a model without covariates the Wald Estimate of the instrumented variable can be interpreted as follows:

$$\beta_{Wtn} = \frac{E[Y_{itn}|Twins_{tn} = 1] - E[Y_{itn}|Twins_{ni} = 0]}{E[C_{itn}|Twins_{tn} = 1] - E[C_{itn}|Twins_{ni} = 0]} \qquad (3.4)$$

The coefficient we obtain from the IV regression thus consists of a numerator that is the difference between the average of the outcome variable Y that we measure for the twinning mothers and the average we measure for the non-twinning mothers. This difference, is weighted by the denominator, which captures the average difference in children between mothers that gave birth to twins at their nth birth and mothers that gave birth to singletons at their nth birth, t years after that nth birth occurred. In the IV model which includes additional covariates (age and age squared in our case) the difference between twinning mothers and non-twinning is adjusted for the difference in children conditional on the covariates, which can be regarded as an even more precise adjustment.
Thus, in contrast to the twins-first methodology the IV regression does account for the fact that the difference in children does not remain equal to 1 as time passes. It also normalizes coefficients so that they consistently reflect the effect of 1 extra child and thus ensures greater comparability of estimates obtained for different times t. What the IV regression can however not correct for, are the effects that the differing probability of twinning and singleton mothers to have young children at home might have on wages and employment. Since these effects are highly dependent

[6]fFor reasons of parsimony and in accordance with other twin studies we refrained from reporting the first stage regression as the descriptive statistics make it more than obvious that twinning significantly increases the number of children that mothers have

on the timing of births they cannot be controlled for by simply adjusting the coefficient in the way the IV estimates do.

Angrist and Imbens (Imbens and Angrist 1994) show that in a setup with heterogeneous effects of the treatment on the outcome variable the IV estimate can be interpreted as the LATE (Local Average Treatment Effect). As Angrist and Schlosser (Angrist et al. 2010b) argue since compliance with treatment is close to perfect in the case of twinning the LATE can be interpreted as the average treatment effect on the non-treated. They advocate an empirical strategy similar to the one we employ where the sample is reduced to mothers that gave at least n births in order to estimate effects of twin birth at the margin.

Angrist and Imbens (Imbens and Angrist 1994) state 3 conditions that need to be fulfilled in order for an instrument to be valid.
The first condition is that the instrument, which in our case is twinning at the nth birth, is correlated with the treated variable, which is the number of children a woman has t years after her nth birth. This condition is definitely fulfilled, even though, as we have shown the strength of the correlation between the number of children and twinning gets weaker as time passes, it nevertheless remains strong and significant for any t and n we consider.
The second condition is monotonicity, this means that the instrument only works in one direction for every treated individual. Again this condition can be assumed to be valid, since the assumption that twinning would somehow lead a mother to reduce the total number of children she had, meaning that after twinning she would have 2 births less than she otherwise would have had seems very hard to justify theoretically and outright contradictory to the story the data tells us.
The final assumption is that there must be no correlation between the instrument and the error term in the regression. This assumption is not formally testable. In general the twinning literature acknowledges three factors that might lead to correlation with the error term. These are the effects of age, race and of in-vitro fertilization on twinning probabilities. The correlation with the error term stems from the fact that these three factors influence at once the twinning probabilities and at the same time might affect our outcome variables. However the problem is very much reduced by the fact that we can directly control for age and race and that most studies still rely on birth-cohorts were IVF did not play an important role.

We argue that the timing of subsequent births is another factor which is bound to result in correlation of the instrument with the error term. The difference in subsequent fertility, or more precisely the fact that, as time passes after birth, the probability of singleton mothers to give birth again and to thus have very young children around them is substantially higher than that of twinning mothers (see fig. 3.2) is going to affect the relative wages and employment we measure. The instrument (twinning) therefore affects the outcome variables (income and employment) not only through the direct variation it induces in the instrumented variable (number of additional children through twinning) but also through another channel (timing of subsequent births). Since timing of subsequent births is thus related to the instrument as well as to the outcome variable it is bound to result in correlation of the instrument with the error term. We therefore have reason to assume that long-term estimates of labor-market outcomes derived with twinning IV models are biased. We can also derive the very probable direction of the induced bias, since it is logical to assume that the wages of the singleton group of mothers are going to be negatively affected by the timing of subsequent births, which gives them a higher probability to have a presence of young children at home as time t passes. We are thus going to underestimate the negative effect of children on a women's career, or β_{Wtn} is bound to be upward biased in our IV estimates as well. For the same arguments as the ones we outlined in our discussion of the twins-first estimates the coefficients we obtain for higher birth parities n should be less less upward biased.

3.7 Results

3.7.1 Main Results

We ran both the twins-first model described by equation 3.2 as well as the IV model described by equation 3.3 and 3.4.This allowed us to see the results that the different methods common in the literature would deliver. We ran our estimations separately for four subsamples of mothers, each corresponding to one of the four birth parities n . A subsample thus always consists of the entirety of mothers giving their nth birth between 1980 and 1992. Thus in order to control for cohort effects we included a series of dummies for the year of birth into the models specified in equation 3.3 and 3.4.

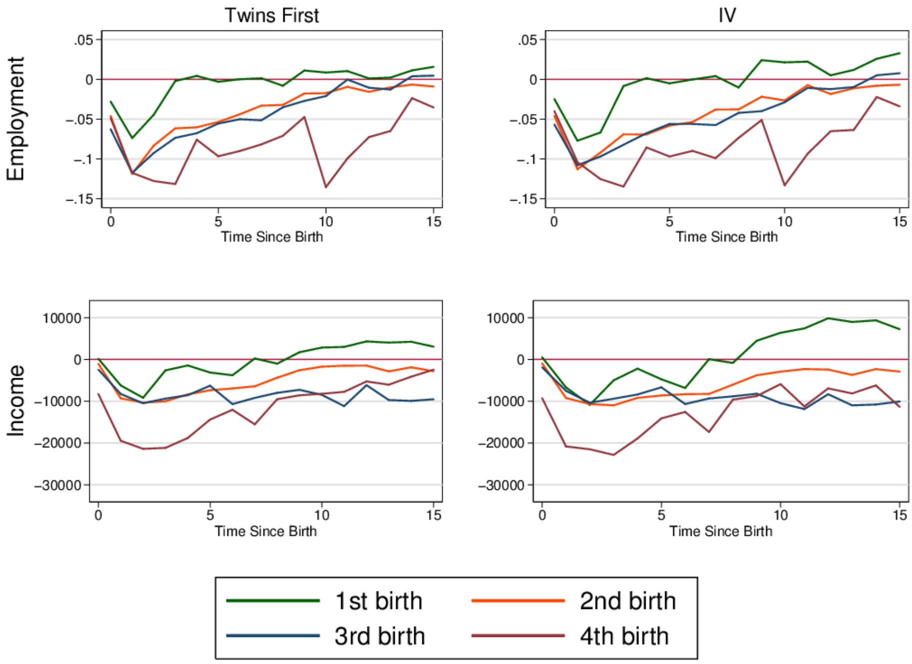

Figure 3.3: Development of Employment for twinnning and non-twinning mothers for selected cohorts

We thus went on to run separate regressions for each of the 2 outcome variables y, labor force participation and gross income, for each of the 4 birth parities n and for the year of birth $t = 0$ as well as the 15 years $t \in (1, 15)$ after birth. This results in a set of $2 \times 4 \times 16 = 148$ regressions for the twins-first, as well as another 148 regressions for the IV estimations. For reasons of parsimony we chose to merely present the parameter of interest graphically here. [7]

In the case of our twins-first model the parameter we are interested in is β_{3tn} from 3.2 and in the case of the of our IV regression it is β_{Wtn} from 3.4. The results of the 296 regressions are shown in Fig. 3.3. The two figures on the left show the results for the twins-first estimations, while those on the right show the estimates from the IV regressions. Every point in a line is the estimated coefficient for a regression set up for time t as denoted on the x-axis and for a parity n which is represented by one of the four different lines. First, we can note that the twins-first and IV estimates behave strikingly similar, as they should, given that the IV-estimate is basically a version of the twins-first estimate that is adjusted for the actual difference in children between twinning and singleton mothers. At the parities where this difference in children remains close to one, it is hard to detect any difference between Twins First and IV estimates. At first birth however, where the difference in children between twinning and singleton mothers goes quite substantially below 1 as time passes, we can see that as we move further down in time, the absolute value of the IV estimate becomes relatively bigger than that of the twins-first coefficient. Again this is what we would expect from 3.4. Since the difference in children between twins and singletons that has dropped significantly below 1, appears in the denominator adjusting the IV coefficient. Basically the IV-coefficients are twins-first coefficients adjusted for the shrinking difference in children in this case. When looking at the individual trajectories a general pattern of a pronounced drop in income and employment for the first two years after birth, followed by a subsequent reduction in the negative effects emerges. In time 0 we are not able to identify the cases in which income or employment were measured prior to birth and those in which it was measured after, which explains the relatively smaller drop in year 0. We can see that the higher the parity n the more pronounced and in particular, the more lasting the effects we find are. We thus find our third hypothesis

[7] tables displaying the full results are available on request

to be entirely confirmed. This is very much in line with what we predicted in the case that bias due to subsequent fertility behavior seriously affects our estimates.

It was not at all theoretically clear that at higher birth orders an additional child would affect a mother more negatively. Stressing such factors as learning and economies of scale in childcare might have led to the opposite conclusion. What did however very accurately predict the results we find is the view that subsequent fertility behavior led our estimates to be heavily upward biased in the case of first birth and does so to a lesser extent for each of the following birth parities.

When looking at the development of the effects of twinning at firstbirth we can see that after about 4 years for employment and after about 7 years for income all negative effects have vanished. Towards the end we actually see some significant positive effects of an extra-child. While not impossible theoretically this is still a rather unexpected result, which does however make perfect sense if, as we outlined before, our estimates in particular for firstbirth, get upward biased with the passing of time. For second and thirdbirth we see negative employment effects that are considerably more pronounced than for firstbirth and which persist for about ten years. While second- and thirdbirth estimates are remarkably similar for employment they do oddly enough diverge quite considerably for income. In both cases negative effects on income persist throughout all 15 years but much more pronouncedly so at third birth. Finally it has to be noted that the estimates for fourth birth, while generally in line with what we would have expected, are considerably less well-behaved than the others and more prone to be easily influenced by outlier values. Overall we have a picture that is remarkably consistent with what we would expect if the biases induced by subsequent fertility behavior were to play a significant role. It has to be noted however that the divergence of second and third birth on income as well as the magnitude of the difference between fourth birth and third birth, given that the fertility behavior at these birth parities does not differ that strongly cannot entirely be explained by simply referring to subsequent fertility behavior. It also has to be noted that the study done by Kahn et al Kahn et al. (2014), which does not employ instrumental variable techniques and thus does not suffer from the potential upward bias we discussed, also finds that effects of having children on maternal labor market outcome become more negative at higher birth parities. However the difference in effect sizes between birth parities that we find tend to be of bigger magnitude than those of the Kahn study, implying that it is relatively likely that the bias of our estimates due to subsequent fertility behavior does at least play some role. The cautious conclusion at this point would be that the circumstantial evidence supports the theory that estimates at firstbirth are downward biased. To shed further light on how plausible it is that the differences are indeed driven by the biases introduced by subsequent fertility behavior, we present additional results on a sample of first-birth mothers over the age of 35 in our section on robustness checks. This sample of firstbirth mothers has a subsequent fertility behavior very similar to that of second and thirdbirth mothers and we also find similar effects on labor market outcomes as we do for the 2nd and third parity. This further strengthens our view that the main driver behind the different effects that we find for different birth parities is that in the case of higher birth parities the instrument becomes "rustier" and our estimates are potentially more upward biased.

Table 3.3: Cumulative Effects on Mothers Income

	Employment years lost			Absolute Income lost			Relative Income Lost		
	after5 yrs	after 10 yrs	after 15 yrs	after 5 yrs	after 10 yrs	after 15 yrs	after 5 yrs	after 10 yrs	after 15 yrs
1st child	-.180	-.141	-.043	-28,965	-25,675	17,235	-.257	-.227	.153
2nd child	-.447	-.624	-.676	-49,586	-78,854	-92,427	-.432	-.686	-.804
3rd child	-.467	-.692	-.711	-44,182	-91,674	-143,692	-.420	-.875	-1.367
4th child	-.586	-1.032	-1.310	-107,484	-161,707	-205,588	-1.193	-1.794	-2.281

In order to gain a better grasp of the magnitude of the effects that an extra child has on labour-market outcomes, table 3.3 shows cumulative effects of fertility variation as estimated by IV. The values displayed are obtained by summing up the IV estimates since time 0 up to the specified time t. Values are provided for years of employment lost, total gross income lost in Danish Kroner, and relative income lost, which is a scaled measure of employment that is obtained by dividing the total income lost by the average income of mothers giving birth at the nth parity in the year before birth was given. We can see that while the employment losses accumulate mostly, shortly after birth, income losses tend to keep on accumulating for a longer time. This is consistent with a story in which women continue pursuing their career after a more serious interruption when the child was very young, but are somewhat set back in their income development due to the time spent out of employment and the continuing constraints that come with having an extra child. Table 3.3 also tells a very clear story about how strongly our estimates differ by birth parity. At first birth the initial negative effects on income and employment are entirely canceled out by later positive effects. If we were to believe the estimates for twinning at first-birth it would follow that having an extra child at first birth results is an accumulative positive employment and income shock (if we assume no discount rate) after 15 years time. On the other hand, for twinning at second birth and higher we do find substantial negative effects, generally close to 1 year of employment and income loss after 15 years due to having had an extra child.

Table 3.4: Regressions for Maternal Employment and Income (firstbirth and secondbirth)

	1st birth			2nd birth		
	0-5 yrs	6-10 yrs	11-15 yrs	0-5 yrs	6-10 yrs	11-15 yrs
Employment						
kids	-0.0349***	0.00353	0.0184	-0.0772***	-0.0429***	-0.0133*
	(-4.45)	(0.28)	(1.37)	(-10.67)	(-5.20)	(-1.70)
age	0.157***	0.164***	0.166***	0.169***	0.176***	0.177***
	(120.74)	(71.44)	(92.60)	(113.65)	(93.17)	(63.72)
age2	-0.00229***	-0.00211***	-0.00193***	-0.00235***	-0.00218***	-0.00200***
	(-103.94)	(-56.40)	(-69.42)	(-98.71)	(-87.95)	(-65.46)
c	-1.892***	-2.311***	-2.716***	-2.113***	-2.546***	-3.003***
	(-114.07)	(-96.25)	(-60.73)	(-75.49)	(-51.89)	(-37.93)
Income						
kids	-4464.9**	-1639.1	9616.7**	-7969.9***	-6419.2***	-1897.8
	(-2.35)	(-0.55)	(2.35)	(-6.28)	(-3.31)	(-0.81)
age	23451.8***	20330.4***	25203.2***	18350.4***	19957.3***	28586.4***
	(70.46)	(36.81)	(47.13)	(56.61)	(42.94)	(34.39)
age2	-310.2***	-225.0***	-245.9***	-226.8***	-214.0***	-286.5***
	(-53.52)	(-24.68)	(-28.68)	(-42.16)	(-33.62)	(-30.55)
c	-246917.9***	-220200.7***	-372810.7***	-166176.5***	-215352.5***	-422696.9***
	(-67.45)	(-39.59)	(-28.51)	(-29.63)	(-18.58)	(-18.05)
N	1416337	1408830	1407581	1135243	1131371	1129316

t statistics in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.5: Regressions for Maternal Employment and Income (thirdbirth and fourthbirth)

	3rd birth			4th birth		
	0-5 yrs	6-10 yrs	11-15 yrs	0-5 yrs	6-10 yrs	11-15 yrs
Employment						
kids	-0.0862***	-0.0497***	-0.00817	-0.104***	-0.0866***	-0.0862***
	(-4.45)	(0.28)	(1.37)	(-10.67)	(-5.20)	(-1.70)
age	0.169***	0.185***	0.200***	0.136***	0.158***	0.174***
	(59.02)	(47.09)	(34.71)	(21.54)	(18.38)	(12.96)
age2	-0.00216***	-0.00216***	-0.00215***	-0.00160***	-0.00176***	-0.00182***
	(-50.51)	(-44.62)	(-35.13)	(-17.58)	(-17.28)	(-13.31)
c	-2.314***	-2.935***	-3.782***	-1.856***	-2.472***	-3.099***
	(-36.06)	(-25.88)	(-21.72)	(-11.47)	(-8.52)	(-6.84)
Income						
kids	-8046.3***	-9450.2***	-10551.0***	-18462.2***	-13399.9**	-6534.6
	(-3.60)	(-3.13)	(-3.14)	(-4.15)	(-2.25)	(-0.84)
age	15239.4***	18747.7***	25999.8***	13937.2***	17845.8***	24566.8***
	(26.33)	(21.88)	(17.61)	(7.20)	(9.11)	(7.17)
age2	-167.0***	-190.0***	-253.1***	-138.1***	-175.4***	-231.5***
	(-18.90)	(-17.31)	(-15.58)	(-4.66)	(-7.22)	(-6.42)
c	-143227.5***	-209966.0***	-371122.1***	-104379.8***	-195752.7***	-383703.4***
	(-12.33)	(-8.98)	(-8.87)	(-2.80)	(-3.37)	(-3.45)
N	380280	378763	377128	85306	84873	84181

t statistics in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

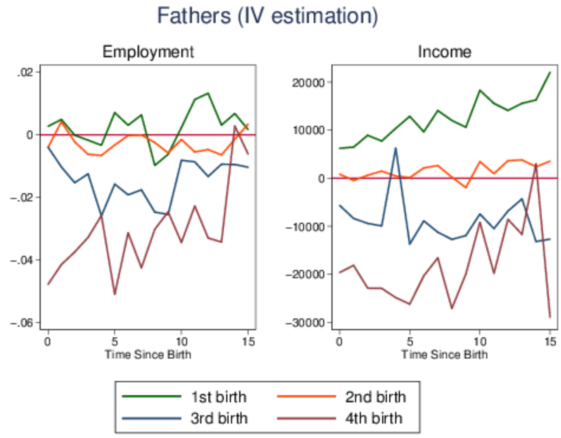

Figure 3.4: Instrumental Variable Estimates for G

Since including the 296 regressions from which the coefficients that are graphed in Fig. 3.3 are taken would probably set a new record for the length of a research paper annex, we decided to instead include a more summarized regression model, which does not estimating yearly coefficients for every t but instead calculates effects for 5 year time periods, meaning the average effect of twinning during year $t = 1-5$, $t = 6-10$, or $t = 11-15$ after birth. Again these models were estimated for all the cohorts giving birth from 1980 to 1992 and thus include dummies for the years at which income was measured, which might in this case range from 1981 (1 year after birth for the 1980 birth cohort) to 2007 (15 years after birth for the 1992 birth cohort). The coefficients for these dummy variables were taken out of the tables. Further since an individual for whom we record values repeatedly might be sampled up to 5 times when looking at effects over 5 year periods our regressions were clustered by individuals.The results of our "summarized" regressions are reported in table 3.5. The story that emerges from them is very similar to what we found when analyzing the graphs.

3.7.2 Additional Results

In addition to our main results on how the effects of fertility shocks affect maternal careers we also took a look at how paternal careers develop in the 15 years after birth. The first interesting fact to note from fig.3.4

is how small the effect on employment is when comparing the coefficient sizes to those of mothers, the income effects are also relatively smaller in particular when accounting for the somewhat higher income that fathers have on average. Even more noteworthy is the very different shape that these curves have compared to the ones we find for mothers. We do not find the initial shock on employment and income that comes right after birth for mothers and is followed by a subsequent recovery . Instead we find lasting effects with relatively little variation over time. We find employment effects that are consistent around zero for first and second birth and slightly negative for third and fourth birth twins. Whereas our paternal income effects turn out to be positive for firstbirth-twinning fathers around zero for second-birth twinning fathers and negative for third- and fourthbirth twinning fathers. These estimates are very consistent with the only other paper in the twinning literature which looks at paternal labor market outcomes. Angrist and Evans (Angrist and Evans 1998) present results for second-birth twins and show that the effects on paternal income and employment are very close to 0 and non-significant. The results we find with fathers are consistent with a story, were an additional child leads to a consistent shift in long-term outlook and behavior of fathers but where the labor-market development is not nearly as substantially interrupted by the intensive care that vey young children require as in the case of mothers. The stylized fact we uncovered that an additional child has a positive effect on paternal earnings at first-birth. But that the effect becomes less positive the further up we move in the birth-order is interesting and demands further research to be fully explained. It also alerts us to the fact that the effects that children have on parental labor-market outcomes, might indeed change with birth-parity, for facts unrelated to subsequent fertility behavior. Thus even though we have good reasons to believe that our maternal results are to an important extent driven by differences in subsequent fertility behavior, the persistent but much smaller differences we find between fathers for different birth-parities alert us to the fact that we should not entirely exclude other explanations for these results.

Finally we were interested in how the effects that children have on maternal careers depend on how much mothers earned before giving birth. From a theoretical standpoint it was not clear what to expect with regard to how effects should vary with income. On the one hand it could be that lower income mothers, due to their relatively greater lack of resources will simply not allow a birth to affect her career to the same extent as a mother

Figure 3.5: Effects of twinning by maternal income

that is better of would and will thus take less time off and risk less of a wage cut. It could also be that the higher paying jobs are more demanding in terms of time investment and that thus mothers who are in top positions suffer relatively more from having a child. On the other hand higher paying jobs often allow employees greater flexibility in accommodating a shock such as child-birth and often offer greater overall job-security which might lead to less of an employment drop among the better paid. Further the greater material and often social resources that are associated with higher pay might make it easier for mothers that are better off to find ways for taking care of their children that do not affect their careers as strongly as those of less well-off mothers.

To uncover the effect of income we looked only the mothers that were in employment the year before birth was given (over 75 % of the mothers in our sample) and then went on to split these mothers into income terciles. We created our terciles based on the entire income distribution of mothers giving birth in a given year. However, since we were interested in relative income status and since our data income data is not deflated, we calculated seperate income distributions for each of the 11 years from 1980 to 1991 (the years preceeding births in 1981 until 1992, we had to exclude 1980 from our estimations here, since we had no information on 1979). Thus the cutoff for which tercile a mother falls in is dependent on her relative position in the income distribution, the year before gave birth. In order to give a sense of the differences between these terciles we note that the average income of mothers falling into the low income tercile was 50,672 DK , while for mothers in the middle tercile it was 114,057DK and 163,205 DK for mothers in the upper tercile in the year before birth was given. Since splitting mothers into income terciles severely reduced our sample size we did not look at births that happened at the fourth parity in our estimations.

Figure 3.5 shows the coefficients we obtained by estimating our IV model described by 3.3 and 3.4 for samples that were split depending on where mothers fall by income tercile. The results show that low and medium income mothers were more severely affected by having an additional child than high income mothers. In particular the initial effect on employment during the first 3 years is more pronounced among lower income mothers. The effects we record for income look relatively similar for mothers in different positions in the income distribution. Keep in mind however that average income in the middle income tercile is twice that in the lower and in the upper tercile it is three times as bigs and you will see that these similar drops in absolute income that we find actually mean that lower income mothers endure a much higher loss in relative income. [8] Another stylized fact that emerges from our estimates is thus that low income mothers are hit more severely in their career development by an additional child than those with higher incomes. This is in accordance with previous results by Budig and Hodges Budig and Hodges (2010)

3.7.3 Robustness Checks

An important assumption of the twinning models we ran is that apart from age and age-squared there is no selection of mothers into twinning.

[8] With the possible exception of thirdbirth were we find quite strong effects for upper income mothers

We tested the extent to which these assumptions were true by running a series of regressions using maternal employment and income in the year before birth was given as outcome variables. If the models are well specified and indeed properly control for any factors selecting into twin-birth, the effect of later twin-birth on previous income or employment should be non-significant once we control for age. These test are not included in most papers looking at twinning, but provide an important test of the endogeneity assumptions, in particular since the number of factors that we know to affect twinning has been growing and since worries about the effects of in-vitro fertilization on twinning estimates are well justified when using more recent data. As we can see in table 3.6 our model is surprisingly enough, not entirely well-specified in controlling for selection in the case of income for first and fourth birth. However the effects we find for twinning are still reasonably small overall.

Table 3.6: Testing for selection

	1st birth		2nd birth		3rd birth		4rth birth	
	Emp	Inc	Emp	Inc	Emp	Inc	Emp	Inc
age	0.165***	27291.1***	0.189***	21868.5***	0.185***	17393.3***	0.152***	16186.9***
	(109.85)	(85.44)	(93.66)	(69.62)	(47.45)	(30.31)	(17.76)	(6.86)
age2	-0.00269***	-397.0***	-0.00287***	-302.7***	-0.00258***	-215.3***	-0.00197***	-184.9***
	(-96.00)	(-63.80)	(-82.62)	(-52.97)	(-41.13)	(-22.43)	(-14.77)	(-4.71)
twin	0.00457	1781.2*	0.00222	191.6	-0.0120	-273.9	-0.0186	-6960.9*
	(0.66)	(1.85)	(0.29)	(0.19)	(-0.82)	(-0.16)	(-0.56)	(-1.79)
_cons	-1.580***	-345892.0***	-2.206***	-295064.1***	-2.536***	-265362.2***	-2.286***	-274622.8***
	(-79.23)	(-86.47)	(-75.87)	(-69.28)	(-42.01)	(-31.19)	(-16.62)	(-7.89)
N	311317	311162	251839	251640	84719	84583	18986	18943

t statistics in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

As discussed, among the different traditionally acknowledged factors that might introduce endogeneity into twin-estimates the one that might be most worrisome in our case is in-vitro-fertilizations. In particular because we observe a big increase of twinning occurrences in our later cohorts. This might be due to later childbirth, but the size of the increase in twinning incidences combined with the fact that in the late 80ies and early 90ies in-vitro fertilization became an accessible technology for the general public make us cautious. Since the endogeneity introduced by a choice variables such as choosing to have an in-vitro fertilization can take on many forms and is hard to predict, we wanted to assure us that our main findings remain valid in a sample where there is little or no endogeneity induced via IVF. We thus re-estimated our IV models using only births that

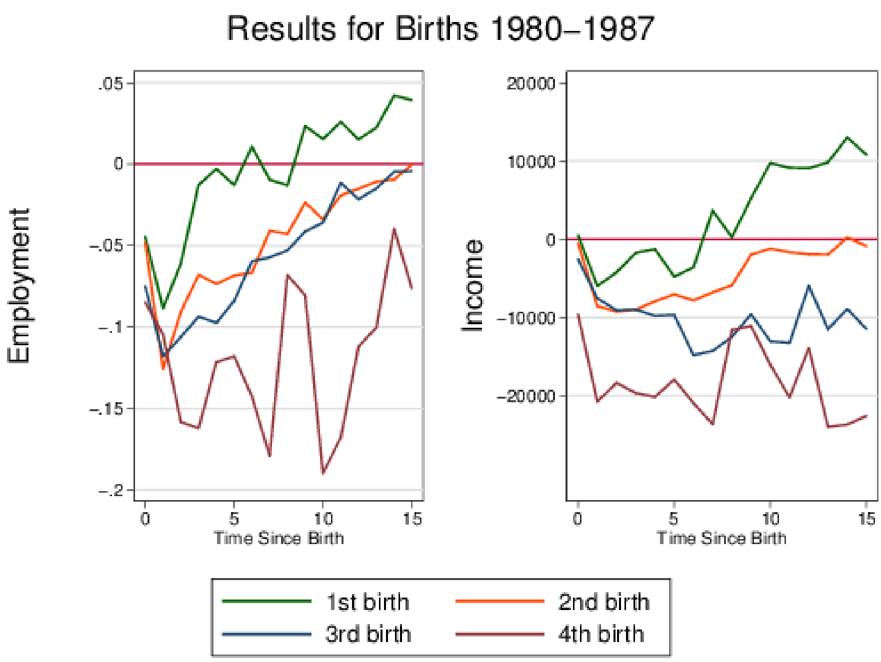

Figure 3.6: Instrumental Variable Estimates for Birth Cohorts 1980 to 1986

happened between 1980 and 1986 when the role of in-vitro-fertilization, was very minor, or basically non-existent. Fig. 3.6 shows our results from those models for our maternal labor market variables.. What we find is very reassuring. The same picture, of increased negative effects of an extra birth on labor market outcomes at higher birth-parities emerges very clearly. Again we find significant positive effects of having an extra child at first-birth after about 7 or 8 years. So all the points that led us to conclude that our estimates were consistent with a story of considerable bias when looking at long-term labor-market outcomes at firstbirth re-emerged. Our results on the development of paternal labor-market incomes as well as on the effects of twinning, dependent on the income distribution also proved robust to this type of specification. [9]

The case we are able to build for our argument that the twinning instrument becomes "less sharp" as time passes and that thus most of our previous estimates are substantially upward biased resides on the one hand on a theoretical argument about how subsequent fertility behavior is bound to affect our estimates and on the other hand on an accumulation of findings, that we would not necessarily expect but that fit well into our theory of upwardly biased estimates. The fact that we find positive coefficients for firstbirth twinning in the long-run and no cumulative negative effect at all is an oddity, that has a good explanation once we accept that the higher probability of singleton mothers to have a very young child at home is bound to negatively affect the wages and employment we record for them at later ts. Also the fact that the effect of an additional child become increasingly negative for higher birth parities n is highly consistent with our view of bias through subsequent fertility behavior. One could easily enough argue that learning by the mothers makes accommodating additional children at higher birthorders easier, rather than harder and that economies of scale allow them to easier cope with an additional child at a higher birth-parity. However if the differences we find are driven by bias due to subsequent fertility our findings are exactly what you would expect. Still it cannot be ruled out that innate differences between the mothers that give more births (or less) are what drives our results or that they are driven by the fact that the economics of the household do in some way change in a way that is more disruptive to maternal careers at higher birth parities. The best test we could come up with for answering these objections was

[9] results available on request

to look at a subsample of firstbirth mothers in which fertility differences between twinning and singleton mothers was much less pronounced, namely mothers giving birth after age 35.

The left graph in Fig. 3.7 shows the subsequent fertility difference between twinning and singleton mothers giving their first birth after age 35 (we denote the sample as >35). It is remarkably similar to the subsequent fertility difference of twinning and singleton mothers after second birth. We can thus make a good case that if the differences in labor market outcomes we found between first- and second birth mothers in our previous estimates were driven by selection into second birth, that then the coefficients we find for our >35 sample should be similar to those we found for other firstbirth mothers. If the different effects on labor market outcomes that we found between our firstbirth and our secondbirth sample were however driven by different subsequent fertility behavior then we would expect our estimates of the >35 sample to be much closer to those we found for the secondbirth sample.

Due to the much smaller sample size in the >35 sample our estimates are relatively noisy as can be seen in the greater volatility of the graphed curves. Nevertheless it is very clear that the effects of twinning on labor-market outcomes are remarkably more negative than those of the normal 1st birth sample. When abstracting from the noise the line that the >35 sample resembles most closely, both in our employment and in our income estimations is that of second birth. This is exactly what we would expect if the differences we previously found between first and secondbirth twinning were the result of bias coming from subsequent fertility. To some extent this also sheds further insight into the question of whether the more negative effects we found at higher birth parities are merely due to mothers being more constrained when they have more children Kahn et al. (2014) or whether part of the difference in the results is due to the fact that our firstbirth estimates suffer from a stronger upward bias. A sample looking only at individuals giving birth over the age of 35 is bound to have severe issues of selection bias attached to it as well and we do not argue that the test we put forward here is conclusive evidence in the form of a mathematical proof. What we do however have is a vast amount of results which all support a story that tells us the traditional estimates we obtained for the effects children have on the labor market outcomes of their mothers were downward biased.

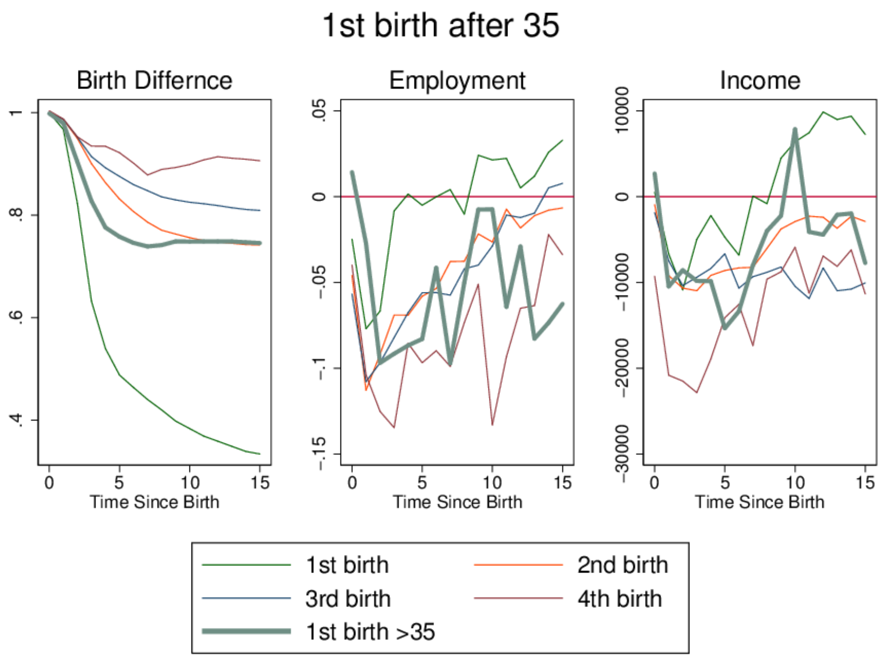

Figure 3.7: Comparing the effects of a firstbirth over the age 35 to full sample estimates for different birth parities

3.8 Conclusion

We revisited the most common approach used to identify the effects of children on a mother's career which is the use of twinning as an exogenous source of variation in the number of children that a mother has. We have shown that when this approach is used to estimate maternal labor market outcomes measured a long-time after birth was given, it might suffer from serious flaws. These flaws arise from the fact that twinning, as well as singleton mothers are bound to differ substantially in their fertility behavior after having given birth for the *n* th time. Since now the wage and employment difference we observe between twinning and singleton mothers is not only a function of the exogenous variation in the number of children that results from twinning but also a function of how subsequent births were on average timed by twinning mothers and by singleton mothers our causal inference is bound to get increasingly biased as time since birth passes and subsequent fertility behavior starts to matter more. We used the term "rusty instruments" to describe the process of the IV identifying strategy becoming increasingly less useful with the passing of time. Since we can identify situations in which the differences in subsequent fertility behavior between twinning and singleton mothers are less pronounced we are able to test the extent to which this bias affects our results. As we show subsequent fertility behavior of twinning and singleton mothers becomes increasingly similar when comparing mothers at higher birth parities or when comparing mothers of higher age. In both cases we consistently find that the negative effects we estimate for an additional child become increasingly bigger. When looking at cumulative effects on employment and income our estimates on a firstbirth sample show that children have as good as no cumulative negative effect over a 15 year time-frame on maternal employment and income. However when looking at higher birth-parity sample we find cumulative effects of an additional child that are close to 1 year in lost employment and income. Since almost all previous studies looking at maternal labor-market outcomes were based on either first- or secondbirth samples we caution to interpret the results showing, generally very small effects of children on mothers career with caution as they might suffer from the upward biases due to subsequent fertility behavior which we outlined. However we must also caution against simply extrapolating the stronger negative effects we find at higher birth parities to the lower ones. The stronger negative effects we find at higher parities are most likely a mix of reduced upward bias and truly stronger effects at higher parities. We also derive additional stylized facts on the effects of children on paternal labor market outcomes. We show that the negative effects of additional children on employment and relative income are bigger for lower income mothers.

When looking at paternal labor market outcomes we found that an additional second child (twinning at first birth) seems to positively affect a fathers income with no effect on employment. But this positive effect becomes consistently less positive when moving to higher birth parities and an additional 4th or 5th child was found to lower paternal employment as well as income. Explaining these stylized facts adequately would require further research.

In addition to contributing to the literature on how children affect maternal labor market outcomes, our work can also be regarded as a cautionary tale about the use of instrumental variable estimating techniques. It goes to show that even when one has found a seemingly perfect instrument leading to clear and plausibly exogenous variation in the treatment variable it is worthwhile to think through the many consequences that an instrumental variable treatment might have. Further we would like to think that embedding the application of instrumental variable estimating techniques into a deep analysis of the context can often times lead to a more insightful reading and interpretation of the encountered effects than a purely mechanical application of the technique.

4 Child Gender and its Effects on Parental Labor Market Participation: A Robust Tale of Danish Parents

Abstract

Many recent studies have stressed that even in industrialized countries the gender of a child can have a significant impact on such parental behaviors as the division of paid work and housework arrangements in the couple as well as of marriage behavior and couple stability. A possible weakness of these studies is that the potential endogeneity of child gender, while acknowledged, is normally not directly accounted for. Using 27 years of registry data including the entire Danish population we first assess which factors influence sex ratios at birth. We then go on to look at whether these factors mediate the effects of child gender on parental labor market outcomes in any meaningful way. We find that the results are remarkably robust and that there is no evidence of mediation. Surprisingly having a son negatively affects the long-term income trajectory of fathers.

4.1 Introduction

Throughout history numerous societies did not only assign very different roles to men and women but also exhibited different preferences towards having sons or daughters. This often implied that the allocation of resources as well as the economic decision-making within households tended to differ strongly, dependent on whether a newborn child turned out to be a son or a daughter. Nowadays such strong child preferences are primarily associated with some societies in South and East Asia. In industrialized countries there is much less evidence of parental discrimination in favor of sons, what we do instead observe seems to be a preference for "balanced" families, with families not having a child of each gender yet, showing a greater propensity to continue having children (Angrist and Evans 1998) (Mills and Begall 2010) . Apart from that, discrimination between boys and girls is mostly viewed as being either relatively small or irrelevant in industrialized countries. For example Meadows shows that child well-being for girls and boys in the United States is indeed very similar (Meadows et al. 2005). However a series of relatively recent studies done by psychologists, sociologists and economists that were made possible either through the availability of big datasets or time-use data have documented that, even though direct gender discrimination is not so much of an issue, child gender still has important effects on family decisions in industrialized nations. Differences dependent on child gender are shown to exist for parenting decisions such as time use, family structure and marital stability. Lundberg (Lundberg 2005b) provides an extensive survey of this literature. The majority of these studies have been done on US data, so as Lundberg (Lundberg 2005b) points out relatively little is known on how strongly results vary across populations or time. Since we have a large longitudinal dataset measuring all births in the Danish population our aim is to partially close that knowledge gap. Another area of concern with the currently existing literature on gender effects is that child gender might not be as exogenous as is often assumed. This is particularly worrisome, since many of the results we have on gendered effects tend to come from studies identifying the existence of small effects using large datasets, thus a small change to the estimate due to omitted variable bias could easily mean that previously significant results were actually insignificant. If some parental characteristic has a small effect on the sex-ratio at birth, but at the same time potentially influences the parental outcomes, which we look at when assessing the effects of child gender our results could easily be driven by ommited variable bias or selection effects. While most studies acknowledge that the sex ratio at birth is not entirely exogenous and might slightly vary due to a great many factors, we are not aware of any that consciously checks the for factors influencing birth probabilities in the sample before going on to estimate effects of child gender. The only factor that the recent economics and sociology literature on child gender consistently controls for, since it is widely acknowledged to affect the probability of child gender is age, and often only maternal age. However the recent biology and demography literature tends to identify a much larger series of factors with the potential to affect birth probabilities. The demographic literature for example includes many examples that across population and ethnic groups the birth probabilities of boys/girls can differ quite significantly (Branum et al. 2009). Further the biological literature has found some evidence for such factors as stress during pregnancy potentially influencing child gender (Hansen et al. 1999). Finally some evolutionary biologists tend to highlight that natural selection might favor species as well as individuals which can adjust the ratio of their offspring (Trivers and Willard 1973), which has led to speculation on factors such as intelligence, height and environmental conditions affecting the sex ratio. These theories remain highly controversial however and have often not stood the test of replication. Further, since child gender at first birth and much more explicitly child sex-composition after second birth influences parental behavior and fertility decisions there is also a line of arguments saying that any births after second birth or potentially even after firstbirth suffer from contamination and selection issues due to the effects of the previous child (Lundberg 2005b). We will go on to discuss these factors and their implications in section 2 and will then continue with discussing the effects of child gender on parental behavior in section 3. In the empirical sections 4 and 5 we first present models exploring factors that influence birth probabilities by looking at the entire Danish population between 1980-92 and then go on to look at the effects of child gender on parental incomes and employment. Section 6 concludes.

4.2 Determinants of Child Gender

While there have been many attempts at and superstitions about how to influence the gender of a new child, ranging from dietary advice to the Chinese Birth Calendar, the generally prevailing view has remained that the gender of a newborn child is practically random. From this perspective, whether a baby turns out to be a boy or a girl represents a "natural experiment" and it has been exploited as such in a significant number of studies in the economics, political science and sociology literature. However the evolutionary biology, medicine and demographic literature are increasingly questioning the extreme version of this point of view, arguing that the probability of the child's gender can indeed be susceptible to a variety of factors.In the following when we refer to the sex-ratio, what we mean is the ratio of males to females in a sample population.

The most commonly acknowledged factors affecting the sex-ratio of children are parental decisions to stop having children depending on sex or to engage in sex-selective abortion in some areas of the world, beyond that there is a more controversial literature exploring so-called "natural factors". For example in the US, the sex ratio, measured as the proportion of males born relative to females, differs quite substantially by ethnicity, also for the US there is a recorded drop in the sex ratio over time in recent years (Branum et al. 2009). A variety of mechanisms might be at play here such as changing preferences and fertility stopping behavior, changing ethnic composition of the United States and changes of the age at which birth is given. Another influential study analyzed the effect of natural factors in a dataset that is remarkably similar to ours, covering the Danish population from 1980-93. It looked at parental age, birthorder and sex of the preceding child, to find that only fathers age had a significant effect on the sex-ratio (Jacobsen et al. 1999)

Trivers and Willard made the point that the ability to adapt the sex-ratio, so that fitter and healthier parents, who are able to invest more in their children have a slightly higher probability of conceiving boys would be evolutionarily beneficial (Trivers and Willard 1973). It has to be noted that this oft-cited theory remains empirically controversial (Brown and Silk 2002). However there is some evidence that not only "natural factors" affect the sex-ratio, but that stressful socioeconomic circumstances can also lower it. Dama (Dama 2011) provides evidence that cross-national variation in sex-ratios is positively influenced by factors directly related to fitness or well-being (and potentially to lower stress) such as a country's GDP and longevity. It has also been shown that the 1959-61 Great Leap Forward Famine in China led to a drop in the sex-ratio followed by a rebound as soon as the famine ended (Song 2012). That economic stressors continue to affect the sex ratio in industrialized nations was shown in a study comparing Eastern and Western Germany after reunification. As the east experienced uncertain economic conditions its sex-ratio fell

significantly relative to that in Western Germany (Catalano 2012). On the micro-level these insights are supported by research showing that women who experience stressful life events have lower sex-ratios (Hansen et al. 1999) as well as by research showing that fathers might also be affected by stress, with consequences such as reduced sperm motility (Fukuda et al. 1996). Theories of stressors affecting sex-ratio also open the possibility of socio-economic status affecting birth-ratios and Teitelbaum and Mantel provide evidence supporting the view that this is the case in the US (Teitelbaum and Mantel 1971). Further, and much less controversial there is a substantial literature on how the gender composition of children might be influenced by "preference" rather than "natural" factors. For example Mills finds that in Europe poverty tends to strengthen a son preference and thus lead to more third births if the first two children were girls Mills and Begall (2010). In general studies treating child gender as a random variable tend to only look at first and second births, since it is known that parental preferences matter substantially for selection into third birth..
While in the absence of "natural factors" first birth can be regarded as random, parental preference might however already start playing a role by selecting parents into or out of continuing to have children dependent on the gender of the first child. Andersson et al Andersson et al. (2006) show that in the Nordic countries, including Denmark, which we study, no evidence for sex preferences mattering at second birth could be found. However in Denmark a girl preference at third birth was documented. Brockmann showed for the case of Germany how sex preferences can change over time Brockmann (2001). It is thus important to note that studies using sex gender at a birth parity higher than 1 to look at parental behavior might already run into self-selection problems of the parents. If the more controversial "natural factor" such as stress affecting the sex ratio, do also play a role selection problems omitted variable problems might already arise at first birth.

It has to be noted that none of the effects influencing the sex-ratio at birth are very big in magnitude. Nevertheless, since many of the effects of child gender on parental decisions are often relatively minor in magnitude as well, we argue that more emphasis has to be placed looking at which factors might possibly influence child gender, in order to avoid that results are driven by omitted variable bias or selection.

Table 4.1: Summary Statistics

Variable Means	Full Sample	1980-86	1986-92
firstbirth			
% boys	.513	.510	.515
maternal age at birth	25.2	24.6	25.8
maternal employment rate in birthyear	.708	.709	.707
maternal monthly gross income in birthyear (DK)	110,832	90,479	131,359
maternal education in years	12.0	11.8	12.2
paternal age	28.3	27.8	28.8
paternal employment rate	.793	.801	.786
paternal monthly gross income in birthyear (DK)	172,845	143,836	202,235
paternal education in years	12.5	12.4	12.7
completed fertility	2.737	2.396	3.081
% married in year of birth	.491	.517	.465
n	340,502	170,977	145,444
second birth			
% boys	.5117	.5101	.5132
maternal age at birth	28.1	27.6	28.5
maternal employment rate in birthyear	.713	.713	.713
maternal monthly gross income in birthyear (DK)	111,880	88,855	135,640
maternal education in years	12.4	12.2	12.5
paternal age	30.9	30.5	31.3
paternal employment rate	.860	.874	.845
paternal monthly gross income in birthyear (DK)	191,358	161,022	223,304
paternal education in years	12.6	12.5	12.8
total number of children	2.695	2.526	2.869
% married in year of birth	.721	.765	.675
n	274,507	139,401	135,106
third birth			
% boys	.5097	.5086	.5108
maternal age at birth	30.8	30.6	31.1
maternal employment rate in birthyear	.713	.713	.713
maternal monthly gross income in birthyear (DK)	101,637	78,287	124,376
maternal education in years	12.4	12.2	12.5
paternal age	33.5	33.2	33.8
paternal employment rate in birthyear	.841	.861	.820
paternal monthly gross income in birthyear (DK)	202,907	173,334	232,341
paternal education in years	12.4	12.2	12.5
total number of children	3.465	3.434	3.582
% married in year of birth	.805	.841	.770
n	92,250	45,514	46,376

4.3 Effects of Child Gender on Parental Behavior

There are different approaches to thinking about the reasons for which a children’s gender might affect parental decisions. The most common explanation brought forward is referred to as a son "preference". Economists would speak of deriving different utilities from the child dependent on its gender. The mechanisms through which child gender might affect utility are open to a variety of interpretations.

A classical way of thinking about the effects of child gender on parental utility is the one brought forward by Ben-Porath and Welch (Ben-Porath and Welch 1976), who discuss children in terms of their potential costs and benefits, dependent on such factors as the ability to provide material support in old-age or the necessity to endow them with bridal gifts etc. Another explanation that is brought forward by economists and sociologists (Lundberg 2005b) is that men have a preference for spending time with boys and thus derive greater utility from investing time in them. Yet another view stresses the production of children as "household goods" and assigns different productivity to parents. This is related to a view, which emphasizes not so much that parents derive different utility from their children, but rather that they face different constraints when investing in them. This view is taken by a strain of literature that emphasizes the relative importance of the presence of fathers in the development of socially and mentally stable boys (see for example Morgan et al (Morgan et al. 1988)). In general it is mostly empirically impossible to distinguish between preference and constraints explanations of the effects we encounter for child gender and it is often reasonable to assume that both are at work. If for example, as has been shown for the United States, a boy leads to a lower parental divorce rate, this could be due to a higher preference of the father for spending time with the boy or due to a relatively higher marginal productivity as measured in how much the father's presence influences the development of a stable child.

Son "preference" can affect such decisions as time spent on caretaking and labor, transfers of goods and consumption decisions, since the positive or negative utility derived from these activities might vary with child gender. In its most extreme form son "preference" can express itself in a different demand for sons and for daughters. Such is the case in some societies in Southeast Asia, such as India, China and formerly Korea were the sex-ratio tends to be heightened by such measures as sex-selective abortion, sex-selective infanticide and differential survival rates of boys and girls (Das Gupta and Shuzhuo 1999). Sen (Sen 1990) famously talked of over "100 million missing women". Das Gupta et. al (Gupta et al. 2003) provide a thorough survey of the reasons for the persistence of this extreme form of son "preference" in some societies.
For industrialized nations the dimensions of parental behavior that have been shown to be influenced by child gender include fertility behavior, marital behavior,and time allocation decisions of which the latter two are closely related to our study.

A series of US studies has shown positive effects of child gender on couple happiness (Barnett and Baruch 1987) (Cox 2003). A study of Danish twins (Kohler et al. 2005) has also found quite substantial happiness increases from having a first-born boy, with those increases being particularly strong among fathers. Further studies done by sociologists in the 1980s found quite strong significant effects of child gender on divorce. For example Morgan et al (Morgan et al. 1988) found a 9% reduction in the risk of marital disruption in the case of boys being born. Diekmann and Schmidtheiny who looked at 16 European countries did also not find any significant effects (Diekmann and Schmidheiny 2004). Morgan and Pollard argue that the effects of gender on divorce tend to decrease over time (Morgan and Pollard 2003) which is in line with newer studies, using large US census samples, such as Bedard and Deschnes (Bedard and Deschênes 2005), who find significant, but much smaller effects. Lundberg and Rose (Lundberg and Rose 2003) also show that the transition rates into marriage of couples who gave birth out-of-wedlock are higher when a boy was born.

In terms of time allocation, several studies have shown that fathers generally spend more time with sons than with daughters (Bryant and Zick 1996) (Yeung et al. 2001). Also fathers have been shown to spend more time on children in general, if one of the them is a boy (Morgan et al. 1988). Lundberg shows that higher father involvement can also be found in couples in which parents are not married (Lundberg et al. 2007). A study on Western German couples (Choi et al. 2005) showed that men increase their work hours in case of having a son or a daughter, but that they do so substantially more in the case of a son. In a separate study done on US data (Lundberg 2005a) it is shown that the reaction of couples to child gender is strongly dependent on education. It finds that boys under 3 increase specializing among lower educated couples, meaning that men work more and women less, while it decreased specialization among higher educated couples. Esping-Andersen and Bonke (Esping-Andersen and Bonke 2007) have looked at parental time use in the Danish case and found increased time investments of fathers if they had boys, with gendered differences being particularly strong among the lower educated.

A fair number of studies in economics and political science acknowledge the fact that child gender might not be entirely endogenous (Lundberg 2005a). To our knowledge none of them does however directly analyze the factors influencing child gender in their sample and then goes on to see whether those factors in particular are driving their results. Lundberg Obviously since we have no perfect model of all the factors driving sex ratio this does not lead to perfect estimates, but rather it would serve as a warning-system of sorts. We could see whether what we know about the endogeneity surrounding child-gender already affects our estimates in a profound way. If estimates change significantly, much more caution would become necessary in interpreting our results, acknowledging that their might be endogenous factors influencing child gender that we are still unaware of and that might have equally profound effects. If however the results are very stable to the factors we are aware off, the probability of other unknown factors profoundly influencing our estimates can also be seen as significantly lower.

4.4 Results on Birth Probabilities

We use the Danish registry data from 1992 to 2007. This Dataset records the entire Danish population. In the version of the registry data that was made available to us people are registered starting at age 15. We can thus retrospectively match children to parents up to 1992. We can also go on to observe the trajectory of parents that gave birth between 1980 and 1992 for another 15 years. One of the big advantages of working with a registry dataset is that worries of sample selection or selective attrition of survey participants do not affect the data. It has to be noted that up to the early 1990s Denmark still remained a remarkably ethnically homogeneous society. The immigrant share of the Danish population was under 3% (under 4% including second-generation immigrants) in 1980, which is the time at which we start looking at our first cohort and under 4 % (under 5% including second-generation immigrants) in 1992 the year in which we start following our last cohort of mothers (Liebig 2007). Unfortunately our data does not include information on whether individuals are immigrants, however the effects should not be too pronounced given their small share of the population for the time of our estimates. Table 4.1 provides a summary of our data. It can be seen that slightly more men than women are born, it can also be seen that the sex-ratio is increasing over time. Interestingly, this is contrary to the trend in the United States where the sex ratio has been decreasing in time (Branum et al. 2009).

Another advantage of our dataset is that it closely resembles that of a prominent study on birth probabilities. Jacobsen et al (Jacobsen et al. 1999) looked at all Danish births happening in the time between 1980 and 1992. That study only looked at demographic factors, namely age of the

Table 4.2: Factors influencing the probabilty of having a boy

	(1) 1st and 2nd births	(2) 1st birth	(3) 2nd birth
maternal age	0.000973	0.00159	0.000389
	(1.06)	(1.26)	(0.29)
paternal age	0.00116^{+}	0.000513	0.00204^{+}
	(1.66)	(0.55)	(1.92)
year	0.00221**	0.00226*	0.00207^{+}
	(3.01)	(2.26)	(1.91)
paternal employment	-0.00241	0.00126	-0.00830
	(-0.26)	(0.10)	(-0.55)
years of education of father	-0.000193	-0.000360	0.000223
	(-0.17)	(-0.23)	(0.14)
maternal employment	-0.00615	0.0000641	-0.0113
	(-0.93)	(0.01)	(-1.17)
maternal education	0.00332*	0.00465*	0.00229
	(2.54)	(2.52)	(1.24)
married at birth	-0.00660	-0.00479	-0.00908
	(-1.11)	(-0.63)	(-0.95)
second birth	0.00520		
	(0.87)		
first child boy			-0.00399
			(-0.49)
_cons	-4.320**	-4.438*	-4.005^{+}
	(-2.96)	(-2.24)	(-1.86)
N	533855	291748	244083

t statistics in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

parents and sex of the previous child to obtain their estimates. The one factor they found to have a significant influence on the gender of children is parental age. Given that several studies on factors influencing sex-ratios have been known to suffer from reproducibility problems, we run models in the same setup as they do. This means we used the age of both parents, as well as the gender of the previous birth (in the second birth sample) as explanatory variables. We did however deviate from the Jakobsen et al. study in two important ways. First, we focused only on first and second births. We did this since our final aim is to evaluate whether studies which use the gender of children as being random might suffer from omitted variable bias due to the effects "natural factors" have on birth probabilites. However at third birth, as was shown for Denmark the sex-composition of previous children affects fertility decisions, leading to the natural effects to be potentially contaminated by "preference" effects. Therefore evaluating children at parities higher than 2 was of no interest to us. Second we added a series of socio-economic variables, namely maternal and paternal employment at birth, the years of education the parents had achieved and whether the couple was married at the time of the birth, to see if those variables also had an influence.

Table 4.3 shows the coefficients from a logit model with child gender at birth as the dependent variable. A child born as a boy is coded as 1 and a girl as 0. The first column shows the results of the sample in which first and second births happening between 1980 and 1992 were pooled together. The second column presents effects on sex-ratio at firstbirth and those in the third column present results for the second birth sample. Reassuringly the results are conform with the findings of the Jacobsen et al study (Jacobsen et al. 1999). They confirm that paternal age tends to be a significant influence in the overall sample as well as the sample on second births, while the other "natural" or demographic factors do not seem to have a significant influence. Among the newly introduced socioeconomic factors, maternal education is found to influence child gender, particularly at first birth. Further, we see a confirmation of the previously noted trend in births, with the birth year of a child having an influence beyond the factors we controlled for.

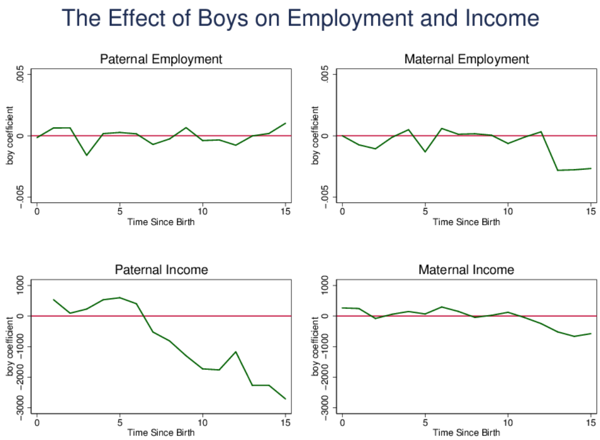

Figure 4.1: Effects of Boys on Parental Labour Market Outcomes

4.4.1 Effects of Gender on Parental Outcomes

We ran several different models in order to estimate the effects of child gender on parental outcomes. The four parental outcome variables Y, which we assessed were maternal and paternal income as well as employment. The basic setup of all our models was the same. We pooled together all the years included in our panel and constructed a new time variable for each pair of parent. This time variable measured the number of years that had passed since birth. If Y_{yi} denotes the value Y takes in a given year y for individual i and by_i denotes the birthyear of the relevant child for individual i then we define a new time variable, time since birth $t = y - by_i$. We then ran a set of seperate regressions pooling together the observations for all individuals with the same t for each regression, to get the average effect of child gender on labout market outcomes t years after birth. The setup of the basic regression which only includes time controls is thus:

$$Y_{t,i} = \beta_{1,t} boy_{by,i} + \beta_{2,t} I_{y,i} + \varepsilon_{t,i} \tag{4.1}$$

Remember that we estimate different betas for different times t that passed since birth, but that we pool individuals having given birth at different years, which have the same amount of years t since birth passed. To control for the time effects this induces $I_{y,i}$ is a set of time dummies that takes the value one if individual i was observed for year y in the sample for time since birth t. These are our basic controls for time effects on income

and employment.
The second model we run also includes the two controls which we have shown to have a significant influence on birth probabilities, namely maternal education and paternal age as measured at birth. It has to be noted that these controls are in no way novel or uncommon when looking at effects on parental decisions. What is novel is to do a stepwise analysis of whether factors that influence the sex ratio, significantly alter the boy or girl effects we find. If they do, this might be an indication that the effects we find for gender might be due to selection effects having to do with the not entirely random nature of child gender.

$$Y_{t,i} = \beta_1 boy_{by,i} + \beta_2 maternaleducation_{by,i} + \beta_3 paternalage_{by,i} + \beta_4 I_{y,i} + \varepsilon t, i \quad (4.2)$$

We finally estimate a third model in which we add all the remaining controls that were also included in the logit estimation on sex ratios. Our time variable t, measuring time since birth could take the value 1 for one individual in 1982 if his child was born in 1981 and also take the value 1 for another individual in 1993 if his child was born in 1992, at which point wages and employment were obviously different which is why the year controls are crucial to the model. We then separately ran those 3 models for our four dependent variables for each of the 15 years after birth which we record $t1 = (1, 15)$. Figure 4.1 shows the boy coefficient for each of the 15 values of t for our four dependent variables in the case of first birth children. What we see is that only one of the four outcome variables we considered, namely paternal income was significantly influenced by the gender of the child. This was also only the case for firstborn children. At second birth the line of regression coefficients remains more or less flat.
Effects on employment of both parents are persistently small and insignificant, as are the effects we observe for maternal income. We can however see that paternal income decreases quite significantly in response to having a boy. Interestingly, this decrease does not happen within the time frame at which studies normally look in order to assess the effects of children on parents. Instead in the first five years, having a boy seems to even have a non-significant but positive effect on parental income. However, when the child is about 8 years old a continuous decline can be observed, which is strongly significant (at $\alpha = 5\%$) for most later years.

As discussed we ran three types of models, including different sets of

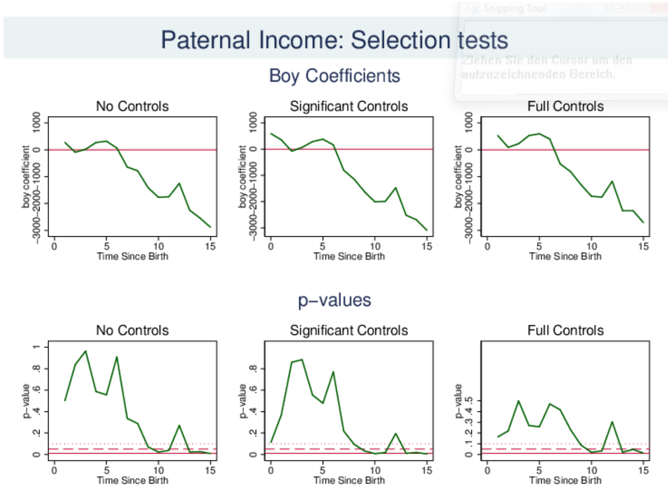

Figure 4.2: Testing for Selection in Paternal Income Regressions

controls. We compared all three models in order to check whether there was any indication that the variables determining birth, which are included in the second model led to a change of the boy coefficients. We also ran Sobel tests, for mediation to test whether, either maternal education or paternal age turned out to be a significant mediator in any of our models. This was not the case. It was remarkable how strongly robust the after all not very large effects (or non-effects) were to the inclusion of variables that do influence the sex-ratio and are bound to also influence the wage development, namely maternal education and paternal age. Fig. 4.2 serves to illustrate how weak and non-significant the mediation of the control variables included in the second and third model was. The graph shows the boy coefficients for our firstbirth sample, by presenting the most interesting case, our regressions on paternal income. The upper 3 graphs in Fig. 4.2 show the set of boy-coefficients we obtained for each of the 15 years after birth. The left graph shows the model with no controls (except year dummies). The second graph shows boy-coefficients in a model in which we control for paternal age and maternal income and the third graph shows boy-coefficients for the model including all of our control variables. In order to be able to assess the significance of these estimates we plotted the p-values that go with each boy coefficient in the graphs below them. When the line drops under the dotted red line the boy coefficient plotted just above is significant at $\alpha = 10\%$, when it drops below the dashed red line, the above coefficient is significant at $\alpha = 5\%$ and when it drops below

the full red line that implies significance at $\alpha = 1\%$. We can see that after about eight years (with the exception of one outlier year) the effects of child gender on parental income turn out significant. We can also see that controlling for factors that we know to affect the sex ratio at birth makes basically no difference to the basic story that emerges. Table 4.3 illustrates our three regression models by reporting coefficients for the case $t = 10$.
As to what factors are exactly driving the decline in male wages that we observe we can only speculate. There is however one potential explanation we can rule out, which is that that the effects we observe might be mediated by marital stability. Earlier studies have shown that boys can increase marital stability (Morgan et al. 1988) and that transition rates into marriage can be higher in the case of having a boy(Lundberg and Rose 2003). This could of course have a mediating effect on wage development. We checked for whether divorce or marriage probabilities were in any way affected by the gender of a newborn child using, simple logit regression models as well as proportional hazard models. In both cases the result was that child gender led to no change in those rates that could be considered significant.

4.5 Conclusion

To some extent this paper was devised under what turned out to probably be a false premise, namely that by not looking cautiously at the factors that influence child-gender, we might find that we are seriously misestimating its consequences. Our results indicate that this is in no way the case. Of course this also has to be taken with caution. Our approach was based on the best current knowledge from the biological and demographic literature on what factors drive sex-ratios, but as determining what drives sex-ratios and their development can in no way be regarded as settled our tests might well be incomplete in important ways. Nevertheless, our results, while probably detrimental to how revolutionary the impact of this paper will be perceived, are reassuring in the sense that they increase confidence in the vast amount of knowledge we have already amassed about the multitude of effects of child-gender at birth.
The paper also has some striking empirical results, namely that only one of the factors we tested seems to be influenced by the gender of the child. Neither maternal income or employment, nor marriage and divorce rates were significantly affected by the gender of the child. What is particularly

Table 4.3: Selection Models for t=10 (year dummies not reported)

	(1)	(2)	(3)
	no controls	basic controls	father1_income
boy	-2883.7**	-3095.9**	-2683.6*
	(-2.61)	(-2.82)	(-2.44)
maternal education (at birth)		20646.2***	10796.0***
		(84.77)	(39.83)
paternal age		921.2***	-2813.6***
		(8.21)	(-19.89)
maternal employment (at birth)			17952.1***
			(13.51)
maternal age			-5426.8***
			(-28.62)
paternal education (at birth)			16959.6***
			(74.14)
marital status (at birth)			21425.4***
			(18.98)
_cons	308540.0***	54621.4***	-175035.0***
	(155.38)	(12.99)	(-29.28)
N	243354	239697	234892

t statistics in parentheses

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

interesting about this finding is that it stands in stark contrast to what Lundberg (Lundberg and Rose 1999) found, which was that male wages tended to rise faster in response to having a son rather than a daughter. The delay in the son effect that is so apparent in our data is also puzzling. First of the most plausible explanation for our results is that fathers are somewhat more likely to reduce hours in the workplace and thus pay in case they have a son at home. This does however need further analysis, preferably using time-use data.

5 Conclusion

This thesis has touched upon two important strains of research in lifecourse studies. First it has looked at how early childhood education can affect the later developmental trajectory of children. Second it has looked at how children, in one case via their existence or birth and in the other via their gender, can affect parental decisions and lifecourses.

The first paper has situated itself in a field of friction in current research on early childhood education and preschool intervention. While most targeted micro-studies find big and lasting effects of early-childhood interventions, the few studies that exist on universal childcare programs have mostly failed to do so. There are several popular explanations for this. One is that small-scale well-designed intervention programs are simply not that well scalable. Another one is that we simply lack the good tools of randomized evaluation that are available when looking at small-scale interventions and are thus simply not able to detect the positive effects of more universal programs. A third explanation is that small-scale interventions target a particular group of vulnerable youths for whom the effects of being put into a good childcare environment are much more beneficial than for the average child. A recent study on the effects of Norwegian childcare has shown some important results that above all seem to support this third argument (Havnes and Mogstad 2010) . It finds that indeed the effects of universal childcare do not seem to be particularly beneficial when calculating the average treatment effect for the entire population. However the authors show that when applying non-linear estimators significant positive effects can be found among more disadvantaged groups.

The aim of my first paper was to test whether this non-linearity property of universal childcare also applies to one of the few other studies that exist on the topic and which looks at the effects of attending high-quality preschool in Denmark. As could be seen the results are mixed. In no way do my results allow to make such a strong case as the Norwegian study and the estimates tend to be relatively close to the mean estimates in most cases. However some important non-linearities in the effects on low-performing boys, who are known to be a particular risk population show that non-linearities, even though weaker ones do also exist in the Danish case. It is important to keep in mind that the counterfactual for high-quality preschool in Denmark is much weaker. We compare high quality preschool to still relatively good lower-quality preschool which might lend some context to the relative weakness of the findings.

The second and third paper of the thesis are concerned with the life-course of parents and how it is affected by their children.
The second paper aims to explore the effects that having a child has on the labor market trajectory of mothers. It does so by applying the most commonly used identification technique used in the literature on this subject namely twinning. The paper does however introduce an important caveat, namely that twinning does not simply imply that one mother has a child more than a potential counterfactual mother. This is because subsequent fertility behavior of twinning and non-twinning mothers differs substantially, non-twinning mothers tend to get more children after giving birth, than their twinning counterparts. It is argued that this is bound to lead to substantial upward bias in the estimates. Several counterfactual groups in which subsequent fertility behavior differs less are constructed and it is shown that indeed in those counterfactual groups, the estimates we obtain for the negative effect of a child on a mothers career are by several orders of magnitude higher.

The third paper looks at another commonly studied theme, namely the effect that a child's gender has on his parents subsequent labor market behavior. The paper draws upon a vast literature in biology and demography, which shows that the probability of having a boy is indeed not entirely random, but influenced by biological as well as socio-economic factors. This key insight is then applied to the economics and sociology literature, which generally acknowledges that some endogeneity might exist but then goes on to treat gender of a child as if it were random.
The paper goes on to reproduce a commonly cited biology paper on factors influencing sex-ratios in Denmark. It adds socio-economic factors to that basic model and comes up with the already previously found result, that paternal age influences birth-probabilities, while adding the result that maternal education also has a significant impact. The paper then goes on to estimate the effects of child gender on parental incomes and labor market outcomes. The only one of those factors that is shown to be seriously influenced by child gender are paternal earnings which tend to drop off

several years after the child was born. All models are run to assess with the controls influencing birth-gender and without them. The results are that as far as we can attempt to measure it omitted variable bias seems to be very low. None of the key results is in any way substantially challenged bu controlling for factors influencing birth probabilities.

Important extensions on all three papers remain for future work.
In Denmark there will soon be a new wave of testscores for the same children that were evaluated in the first paper available. This would allow to assess the longer-term effects of preschool. Do the uncovered weak non-linearities persist? Do they grow stronger or do they vanish as time passes?

The second paper assesses the most common instrument for looking at an exogenous variation in the number of children. However there are other prominent instruments, in particular the gender composition of the first two-children has been used. Since parents tend to prefer mixed-birth children, parents with two sons or two daughters are more likely to get children. This instrument might suffer from equal problems of timing as well as from selection problems, since a mixed-gender preference might not be uniformly distributed across the population. Would equal changes to the estimates occur when attempting to control for those factors?

The third paper, shows some surprising results about paternal income in response to a boy. However those results are hard to interpret without clear information on why they earn less. Do they go for less stressful careers ? Do they reduce working hours? These are questions that could be answered using the Danish time-use data to gain a more complete picture.

What unites these papers is on the one hand their common focus on how a certain life-course event affects later outcomes and on the other hand a common underlying methodological philosophy. The title of the thesis "Cautious Inference", is at once meant as an embrace and as a critique of the empirical design revolution that has swept through the social sciences. We do nowadays have a much better toolkit to derive formal causal estimates. The availability of this toolkit is however sometimes to easily treated as the sole purpose of research. Identifying a source of exogenous variation becomes the main focus of a study and the variables are simply thrown into the "causality grinder" to get an outcome. The point of all three papers has been in a way to take the causal approach and try to add some circumstantial perspective. Most of the time, existing results are strengthened by this approach, which is valuable as such. And in some cases our results and their interpretation can even change entirely through cautious thinking about the many ways in which exogenous variation can affect our results.

Bibliography

Abadie, A., J. Angrist, and G. Imbens (2002, January). Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica 70*(1), 91–117.

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

Andersson, G., K. Hank, M. RÃ¸nsen, and A. Vikat (2006). Gendering family composition: Sex preferences for children and childbearing behavior in the nordic countries. *Demography 43*(2), 255–267.

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

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

Angrist, J. D. (1991, February). Grouped-data estimation and testing in simple labor-supply models. *Journal of Econometrics 47*(2-3), 243–266.

Angrist, J. D. and W. N. Evans (1998, June). Children and their parents' labor supply: Evidence from exogenous variation in family size. *American Economic Review 88*(3), 450–77.

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.

Baker, M., J. Gruber, and K. Milligan (2008). Universal child care, maternal labor supply, and family well-being. *Journal of Political Economy 116*(4), pp. 709–745.

Barnett, R. C. and G. K. Baruch (1987). Determinants of fathers' participation in family work. *Journal of Marriage and Family 49*(1), pp. 29–40.

Becker, G. S. (1994, July). *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education (3rd Edition)*. Number beck94-1 in NBER Books. National Bureau of Economic Research, Inc.

Becker, G. S. and H. G. Lewis (1973, Part II,). On the interaction between the quantity and quality of children. *Journal of Political Economy 81*(2), S279–88.

Becker, G. S. and N. Tomes (1976, August). Child endowments and the quantity and quality of children. *Journal of Political Economy 84*(4), S143–62.

Bedard, K. and O. Deschênes (2005). Sex preferences, marital dissolution, and the economic status of women. *Journal of Human Resources 40*(2).

Ben-Porath, Y. and F. Welch (1976). Do sex preferences really matter? *The Quarterly Journal of Economics 90*(2), pp. 285–307.

Berlinski, S., S. Galiani, and M. Manacorda (2008, June). Giving children a better start: Preschool attendance and school-age profiles. *Journal of Public Economics 92*(5-6), 1416–1440.

Black, S., P. Devereux, and K. Salvanes (2005, June). The more the merrier? the effect of family size and birth order on childrens education. CEE Discussion Papers 0050, Centre for the Economics of Education, LSE.

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. *The Quarterly Journal of Economics 122*(1), 409–439.

Branum, A. M., J. D. Parker, and K. C. Schoendorf (2009). Trends in us sex ratio by plurality, gestational age and race/ethnicity. *Human Reproduction 24*(11), 2936–2944.

Brockmann, H. (2001). Girls preferred? changing patterns of sex preferences in the two german states. *European Sociological Review 17*(2), 189–202.

Brown, G. R. and J. B. Silk (2002). Reconsidering the null hypothesis: Is maternal rank associated with birth sex ratios in primate groups? *Proceedings of the National Academy of Sciences 99*(17), 11252–11255.

Browning, M. (1992, September). Children and household economic behavior. *Journal of Economic Literature 30*(3), 1434–75.

Bryant, W. K. and C. D. Zick (1996). An examination of parent-child shared time. *Journal of Marriage and Family 58*(1), pp. 227–237.

Buchmann, C., T. A. DiPrete, and A. McDaniel (2008). Gender inequalities in education. *Annual Review of Sociology 34*(1), 319–337.

Budig, M. J. and M. J. Hodges (2010). Differences in disadvantage: Variation in the motherhood penalty across white women's earnings distribution. *American Sociological Review 75*(5), 705–728.

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

Carrasco, R. (1998). Binary choice with binary endogenous regressors in panel data: Estimating the effect of fertility on female labour participation. Papers 9805, Centro de Estudios Monetarios Y Financieros-.

Catalano, R. (2012). Sex ratios in the two germanies: a test of the economic stress hypothesis. *Human Reproduction 18*(9), 1972–1975.

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

Choi, H.-J., J. M. Joesch, and S. Lundberg (2005, September). Work and family: Marriage, children, child gender and the work hours and earnings of west german men. IZA Discussion Papers 1761, Institute for the Study of Labor (IZA).

Cox, D. (2003). Private transfers within the family: Mothers, fathers, sons and daughters. Boston College Working Papers in Economics 605, Boston College Department of Economics.

Cramer, J. a. C. (1980). Fertility and female employment: Problems of causal direction. *American Sociological Review 45*(2), pp. 167–190.

Currie, J. (2001, Spring). Early childhood education programs. *Journal of Economic Perspectives 15*(2), 213–238.

D, F. M. (2008, November). Starting school at four: The effect of universal pre-kindergarten on children's academic achievement. *The B.E. Journal of Economic Analysis & Policy 8*(1), 1–40.

D'Addio, A. C. and M. M. d'Ercole (2005, September). Trends and determinants of fertility rates: The role of policies. OECD Social, Employment and Migration Working Papers 27, OECD Publishing.

Dama, M. S. (2011, 08). Sex ratio at birth and mortality rates are negatively related in humans. *PLoS ONE 6*(8), e23792.

Das Gupta, M. and L. Shuzhuo (1999). Gender bias in china, south korea and india 1920-1990: Effects of war, famine and fertility decline. *Development and Change 30*(3), 619–652.

Datta Gupta, N. and M. Simonsen (2010, February). Non-cognitive child outcomes and universal high quality child care. *Journal of Public Economics 94*(1-2), 30–43.

Datta Gupta, N. and M. Simonsen (2011, July). Where to put the kids? effects of type of non-parental child care on pre-teen skills and risky behavior. (5848).

Diekmann, A. and K. Schmidheiny (2004). Do parents of girls have a higher risk of divorce? an eighteen-country study. *Journal of Marriage and Family 66*(3), 651–660.

Esping-Andersen, G. 2. A Child-Centred Social Investment Strategy. *Why We Need a New Welfare State*, 26–68.

Esping-Andersen, G. and J. Bonke (2007). Parental investments in children : how bargaining and educational homogamy affect time allocation. *Demosoc Working Paper 20*.

Esping-Andersen, G., I. Garfinkel, W.-J. Han, K. Magnuson, S. Wagner, and J. Waldfogel (2012). Child care and school performance in denmark and the united states. *Children and Youth Services Review 34*(3), 576 – 589. <ce:title>Comparative Child and Family Policy</ce:title>.

Feyrer, J., B. Sacerdote, and A. D. Stern (2008). Will the stork return to europe and japan? understanding fertility within developed nations. *The Journal of Economic Perspectives 22*(3).

Frölich, M. and B. Melly (2008, January). Unconditional quantile treatment effects under endogeneity. IZA Discussion Papers 3288, Institute for the Study of Labor (IZA).

Fukuda, M.and Fukuda, K., T. Shimizu, W. Yomura, and S. Shimizu (1996). Kobe earthquake and reduced sperm motility. *Human Reproduction 11*, 1244–1246.

Goldin, C. (1995, July). Career and family: College women look to the past. Working Paper 5188, National Bureau of Economic Research.

Goodman, R. (1997). The strengths and difficulties questionnaire: a research note. *The Journal of Child Psychology and Psychiatry and Allied Disciplines 38*(5), 581–586.

Grogger, J. and S. G. Bronars (2001, June). The effect of welfare payments on the marriage and fertility behavior of unwed mothers: Results from a twins experiment. *Journal of Political Economy 109*(3), 529–545.

Gronau, R. (1974, Jan-Jun). The effect of children on the housewife's value of time. In *Economics of the Family: Marriage, Children, and Human Capital*, NBER Chapters, pp. 457–490. National Bureau of Economic Research, Inc.

Gupta, M. D., J. Zhenghua, L. Bohua, X. Zhenming, W. Chung, and B. Hwa-Ok (2003). Why is son preference so persistent in east and south asia? a cross-country study of china, india and the republic of korea. *The Journal of Development Studies 40*(2), 153–187.

Hansen, D., H. Møller, and J. Olsen (1999). Severe periconceptional life events and the sex ratio in offspring: follow up study based on five national registers. *BMJ 319*, 548–549.

Hardin, J., S. L. Carmichael, S. Selvin, and G. M. Shaw (2009, 1). Trends in the probability of twins and males in california, 1983 to 2003. *Twin Research and Human Genetics 12*, 93–102.

Havnes, T. and M. Mogstad (2010, May). Is universal child care leveling the playing field? evidence from non-linear difference-in-differences. IZA Discussion Papers 4978, Institute for the Study of Labor (IZA).

Havnes, T. and M. Mogstad (2011, May). No child left behind: Subsidized child care and children's long-run outcomes. *American Economic Journal: Economic Policy 3*(2), 97–129.

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science 312*(5782), 1900–1902.

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

Herbst, C. M. and E. Tekin (2010, August). Child care subsidies and child development. *Economics of Education Review 29*(4), 618–638.

Hoekstra, C., Z. Zhao, C. Lambalk, G. Willemsen, N. Martin, D. Boomsma, and G. Montgomery (2008). Dizygotic twinning. *Human Reproduction Update 14*(1), 37–47.

Imbens, G. W. and J. D. Angrist (1994). Identification and estimation of local average treatment effects. *Econometrica 62*(2), pp. 467–475.

Jacob, B. A. (2002, May). Where the boys aren't: Non-cognitive skills, returns to school and the gender gap in higher education. NBER Working Papers 8964, National Bureau of Economic Research, Inc.

Jacobsen, J. P., J. W. P. III, and J. L. Rosenbloom (1999). The effects of childbearing on married women's labor supply and earnings: Using twin births as a natural experiment. *Journal of Human Resources 34*(3), 449–474.

Jacobsen, R., H. Møller, and A. Mouritsen (1999). Natural variation in the human sex ratio. *Human Reproduction 14*(12), 3120–3125.

Jaumotte, F. (2003a). Female labour force participation: Past trends and main determinants in oecd countries. OECD Economics Department Working Papers 376, OECD Publishing.

Jaumotte, F. (2003b, December). Female labour force participation: Past trends and main determinants in oecd countries. OECD Economics Department Working Papers 376, OECD Publishing.

Kahn, J. R., J. GarcÃa-Manglano, and S. M. Bianchi (2014). The motherhood penalty at midlife: Long-term effects of children on women's careers. *Journal of Marriage and Family 76*(1), 56–72.

Katz, L. C. and C. J. Shatz (1996, November). Synaptic activity and the construction of cortical circuits. *Science (New York, N.Y.) 274*(5290), 1133–1138.

Knudsen, E., J. J. Heckman, J. Cameron, and J. P. Shonkoff (2006, July). Economic, neurobiological and behavioral perspectives on building america's future workforce. *World Economics 7*(3), 17–41.

Knudsen, E. I. (2004, October). Sensitive periods in the development of the brain and behavior. *J. Cognitive Neuroscience 16*, 1412–1425.

Kohler, H.-P., J. R. Behrman, and A. Skytthe (2005). Partner + children = happiness? the effects of partnerships and fertility on well-being. *Population and Development Review 31*(3), pp. 407–445.

Leamer, E. E. (1983, March). Let's take the con out of econometrics. *American Economic Review 73*(1), 31–43.

Leamer, E. E. (2010). Tantalus on the road to asymptopia. *Journal of Economic Perspectives 24*(2), 31–46.

Lefebvre, P., P. Merrigan, and M. Verstraete (2008). Childcare policy and cognitive outcomes of children: Results from a large scale quasi-experiment on universal childcare in canada. Cahiers de recherche 0823, CIRPEE.

Liebig, T. (2007). The labour market integration of immigrants in denmark. OECD Social, Employment and Migration Working Papers 50, OECD Publishing.

Loeb, S., M. Bridges, D. Bassok, B. Fuller, and R. W. Rumberger (2007, February). How much is too much? the influence of preschool centers on children's social and cognitive development. *Economics of Education Review 26*(1), 52–66.

Lopez, N. (2003). *Hopeful Girls, Troubled Boys: Race and Gender Disparity in Urban Education*. Routledge.

Lundberg, S. (2005a, September). The division of labor by new parents: Does child gender matter? IZA Discussion Papers 1787, Institute for the Study of Labor (IZA).

Lundberg, S. (2005b). Sons, daughters, and parental behaviour. *Oxford Review of Economic Policy 21*(3), 340–356.

Lundberg, S., S. McLanahan, and E. Rose (2007). Child gender and father involvement in fragile families. *Demography 44*(1), pp. 79–92.

Lundberg, S. and E. Rose (1999, October). The effect of sons and daughters on men's labor supply and wages. Working Papers 0033, University of Washington, Department of Economics.

Lundberg, S. and E. Rose (2003). Child gender and the transition to marriage. *Demography 40*(2), pp. 333–349.

Magnuson, K. A., C. Ruhm, and J. Waldfogel (2007, February). Does prekindergarten improve school preparation and performance? *Economics of Education Review 26*(1), 33–51.

Meadows, S. O., K. C. Land, and V. L. Lamb (2005). Assessing giligan vs. sommers: Gender-specific trends in child and youth well-being in the united states, 1985-2001. *Journal of Labor Economics 70*, 1–52.

Mills, M. and K. Begall (2010). Preferences for the sex-composition of children in europe: A multilevel examination of its effect on progression to a third child. *Population Studies 64*(1), 77–95. PMID: 20391202.

Mincer, J. (1963). *Opportunity costs and income effects*. Stanford, CA: Stanford University Press.

Morgan, S. and M. Pollard (2003). Do parents of girls have a higher risk of divorce? revisiting father involvement and marital stability. *Working Paper*.

Morgan, S. P., D. N. Lye, and G. A. Condran (1988). Sons, daughters, and the risk of marital disruption. *American Journal of Sociology 94*(1), pp. 110–129.

OECD (2011). Oecd family database. Technical report, OECD, Paris.

OECD (2012). Oecd family database.

Rindfuss, R., D. Guilkey, S. Morgan, y. Kravdal, and K. Guzzo (2007). Child care availability and first-birth timing in norway. *Demography 44*(2), 345–372.

Rosenzweig, M. R. and K. I. Wolpin (1980, January). Testing the quantity-quality fertility model: The use of twins as a natural experiment. *Econometrica 48*(1), 227–40.

Schultz, T. (1977). *The influence of fertility on the labor force behavior of married women*. Technical analysis paper. available from National Technical Information Service.

Sen, A. (1990). More than 100 million women are missing. *New York Review of Books 37*(20), 61–66.

Shonkoff, J. P. and D. A. Phillips (2000). *From Neurons to Neighborhoods: The Science of Early Childhood Development*. The National Academies Press.

Simonsen, M. (2010, 09). Price of high-quality daycare and female employment. *Scandinavian Journal of Economics 112*(3), 570–594.

Smith, J. P. and M. P. Ward (1985). Time-series growth in the female labor force. *Journal of Labor Economics 3*(1), pp. S59–S90.

SMITS, J. and C. MONDEN (2011). Twinning across the developing world. *PLoS ONE 6*(9).

Song, S. (2012). Does famine influence sex ratio at birth? evidence from the 1959-1961 great leap forward famine in china. *Proceedings of the Royal Society B: Biological Sciences 279*(1739), 2883–2890.

Teitelbaum, M. S. and N. Mantel (1971, 1). Socio-economic factors and the sex ratio at birth. *Journal of Biosocial Science 3*, 23–42.

Trivers, R. L. and D. E. Willard (1973). Natural selection of parental ability to vary the sex ratio of offspring. *Science 179*(4068), 90–92.

Waldfogel, J. (1998). Understanding the "family gap" in pay for women with children. *The Journal of Economic Perspectives 12*(1), pp. 137–156.

Westergaard, T., J. Wohlfahrt, P. Aaby, and M. Melbye (1997, 3). Population based study of rates of multiple pregnancies in denmark, 1980-94. *BMJ 314*(7083), 775.

Wilde, E. T., L. Batchelder, and D. T. Ellwood (2010, December). The mommy track divides: The impact of childbearing on wages of women of differing skill levels. Working Paper 16582, National Bureau of Economic Research.

Yeung, W. J., J. F. Sandberg, P. E. Davis-Kean, and S. L. Hofferth (2001). Children's time with fathers in intact families. *Journal of Marriage and Family 63*(1), 136–154.