Rusty Instruments? Revisitting the Twin Approach to Estimating the Relationship between Fertility and Maternal Labour Market Outcomes

Sander Wagner

July 2013

Abstract

This paper revisits the link between fertility and subsequent maternal labour-market outcomes. Using Panel-Data for the entire Danish population we are able to present precise estimates, showing the effects of having a child for 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. By not properly taking into account differing subsequent fertility behavior of twinning and singleton mothers estimates that are derived for long-term outcomes are bound to be upward biased. We show that this worsening of estimates is less of a problem when analyzing twinning at higher birthorders. The effects of children on labor-market outcomes are bigger and more lasting when looking at a higher birthorder sample, cumulatively resulting in the equivalent to about a year of lost labour in 15 years. We build a case that this difference is due to the fact that the instrument loses more sharpness at lower birth parities. We also present new results on the effect of children on paternal outcomes and show that lower income mothers are more heavily affected in their labour market trajectory.

1 Introduction

Over the last century industrialized nations have experienced a vast increase in the supply of female labour, as well as a strong decline in fertility [31]. These developments are generally regarded as related. However, pinning down the exact relationship between labour-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 difficult to circumvent the underlying endogeneity problems and to know more clearly how factors, such as labour-market participation and wages affect fertility and vice versa. While both, the effect of labour-market outcomes on fertility and inversely the effect of fertility on a mothers' labour-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 dataset of the entire Danish population, we are able to show on a year-by-year basis how fertility shocks develop over time. More precisely we look at the effects of fertility on labour-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, such as characteristics related to self-selection into having more children or a change in the household economy due to the presence of more children. In this scenario it would be differences between households tgat explain why we encounter so strongly contrasting effects at different birth parities. The second 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 birthorders. 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 behaviour. 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 labour 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 differing subsequent fertility behaviour 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 a shrinking difference of children between twinning and non-twinning mothers is invariably linked to that 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 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 behaviour 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 one 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 effects that subsequent fertility behaviour has on our estimates becomes increasingly less problematic. Our estimates show that previous results in the literature on the relationship between fertility and maternal labour market outcomes probably underestimate the depth and the duration of the negative shock for a mothers carreer that results 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 of firstbirth [29] [24] or of secondbirth twins[5], were naturally you tend to have a much a greater sample size than at higher parities . Another weakness of the vast majority of previous estimates is that they are done almost exclusively with cross-sectional data[2], which does not allow them to distinguish between cohort effects and time effects properly, when looking at outcomes that lie at different points in time after a childs birth. The panel nature of our data as well as the big sample size it offers allows 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 literature on uncovering the a relationship between fertility and labour-market outcomes as well as on discovering sources of exogenous variation in fertility. We then go on to describe our data as well as the situation of mothers in Denmark. This is followed by a discussion of the different empirical strategies we employ to look at the effects of twin-birth on fertility as well as by a discussion of the importance of accounting for birth-order. Then the main results are presented, followed by additional results and a series of checks on their robustness. Finally, we conclude by placing

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

[2]A notable exception is the use of panel data by Carrasco [11]

our findings into the context of the previous literature.

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 labour-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 labour-market integration rates 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 labour market participation, that are suddenly displaying much higher fertility rates than more traditionalist societies, like Italy or Spain [17]. While these broad macro-trends may show a reversal in the relationship between a countries' female labour-force participation and fertility, at the micro-level, the historical, theoretical and empirical evidence still suggests that if you look at the labour-market trajectory of an individual mother, there is a negative relationship between fertility rates and female labour-force participation.

This crude association is supported by a series of detailed historical studies. For example Goldin [18] 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 to affect a mothers labour-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 [20] argues that the dominant labour-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 labour-market participation and income. Waldfogel provides a thorough survey of this literature [33]. However, as Browning noted in his review on children and the economic behaviour 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 labour-market outcomes in a satisfactory way. Thus one had to be cautious about drawing inferences [9].

As Angrist noted, nothing illustrates the inherent endogeneity problem more clearly than the fact that economists run regressions with labour-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 labour-market characteristics [5][3].

Since Browning's critique, there has been an increasing number of studies aiming to look at the relationship between fertility and labour market outcomes in setups that allow for causal inference.

The attempt to achieve valid causal inference has mainly been done via instrumental variable estimation techniques. However the search for variables that correlate with fertility but have no effect on labour-market outcomes, also known as valid instruments, 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 [13] as well as the country of origin [30]. 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 labour supply [29]. 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 [19]. 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 [24]. Angrist and Evans [5] 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. Thus, 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. 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 Labour Force Participation using those as an instrument [1]. 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 [12].

These studies have overwhelmingly found that fertility shocks have a negative effect on mothers labour

[3]A notable exception was Mincer who insisted on not including fertility variables in labour market outcome regressions [27]

market outcomes, but that the effect is much smaller than a standard OLS estimation would imply[4]. Also, effects have found to be nonpersistent. The estimates on when the effects of an extra child perish differ, but they tend to range between 2 and 13 years.

A second line of research in the twinning literature has not focused on maternal outcomes but instead looked at the effects that fertility shocks have on child outcomes. These studies have predominantly tested the quantity-quality model of children going back to Gary Becker and Lewis[6] as well as to Becker and Thomes [7]. This Literature includes work by Black, Devreux and Salvanes looking at effects of additional siblings on children's educational attainments in Norway [8], Caceres looking at school outcomes [10] and Angrist, Levy and Schlosser Testing a series of human capital related outcomes such as earnings and education in Israel. [2]

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 at a given point in the birthorder. Good examples include the work by Black, Devreux and Salvanes [8] as well as that by Angrist, Levy and Schlosser [8], 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 [12] on maternal health and well-being to actually analyze the effects at the margins. The entire previous literature maternal labour-market outcomes normally just looks at twinning at first birth [29] [24] [19] or at second birth [5] [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 [33] 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.

Our paper falls firmly into the camp of the literature looking at the effects of a variation in the number children 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

[4] An exception is the Aguero and Marks study which finds no effects [1]

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

Birthtype	Frequency	Percentage
Singletons	1,845,945	99.02
Twins	18,106	.097
Triplet	214	.01
Quadruplets	11	.00
Quintuplets	3	.00
Sixtuplets	2	.00
Total	1,864,281	100.00

Table 1: All Danish Births 1980-92

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 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 in which we start following our last cohort of mothers [26]. Add to that that over half of the immigrants living in Denmark at the time were from other OECD countries [26]. 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 [14], as can be seen in the small differences in total realized fertility between our 1980 and 1992 birthcohorts (see Tab. 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 [15]

It has to be noted that Denmark has extremely generous legislation for assisting mothers [28]. There are 18 weeks of paid maternity leave. Further DKK 8,024 a month is paid for the second and then for each subsequent child. 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, since the same sum is paid for every additional child as of the second child. This is important since a major part of our conclusions depends on analyzing the different behavioral response of mothers to an additional birth at different points in the birthorder. 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 [16], 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 [25].

The Danish Registry records key demographic and economic variables for the entire population on a yearly basis. Every person enters the registry data at age 15. While the number of children that a family has are recorded according to their age group 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 birthdate 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 twinbirths 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 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 absolute numbers, in our firstbirth sample than in the second or thirdbirth 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. 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 $t \in [0, 15]$ to capture the time that has passed since birth was given. 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 birthyears in out models to control for time effects.

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 [22] [32]. 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 [34], 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

	1st birth		2nd birth		3rd birth		4th birth	
	twins	*singletons*	*twins*	*singletons*	*twins*	*singletons*	*twins*	*singletons*
Total no. 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
Maternal 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
Maternal Education in Years								
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
Maternal Income Year Before 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
Maternal Employment Year Before 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
Paternal Income Year Before 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
Paternal Employment Year Before 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 2: Mean of Mother Characteristics Recorded at Year of Birth

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 [21] and is generally acknowledged among economists and demographers. What has received far less attention is how the adaption in fertility behaviour 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 behaviour (see for example citeangrist10. More precisely it is generally pointed out in the descriptive statistics on total amount of children that twinning and non-twinning mothers have that 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 behaviour is bound to affect estimates remains 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.

4.1 The effects of twinning on subsequent fertility

The interest of this paper and of the twinning literature in general is to derive estimates about the effects of fertility on long-term labour 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 labour 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

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

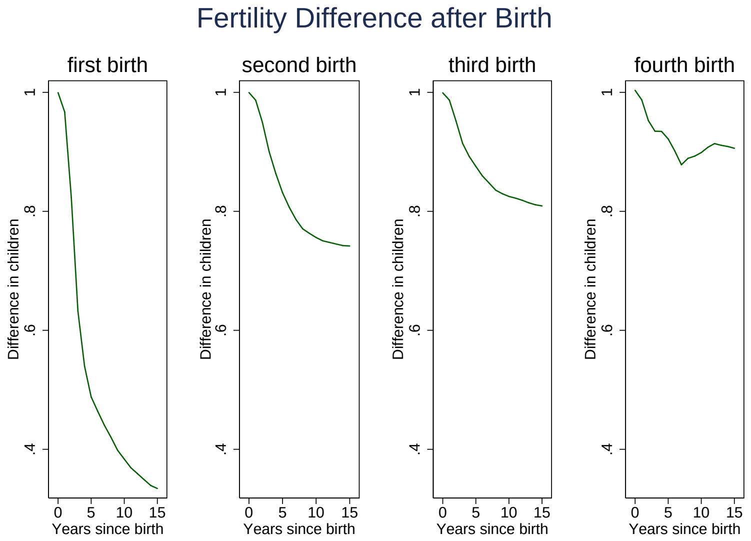

Figure 1: Difference in Fertility Between Twin and Non-Twin Mothers

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 showed 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 difference 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 [15]. In order to illustrate the logic of our four assumptions 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 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 sin-

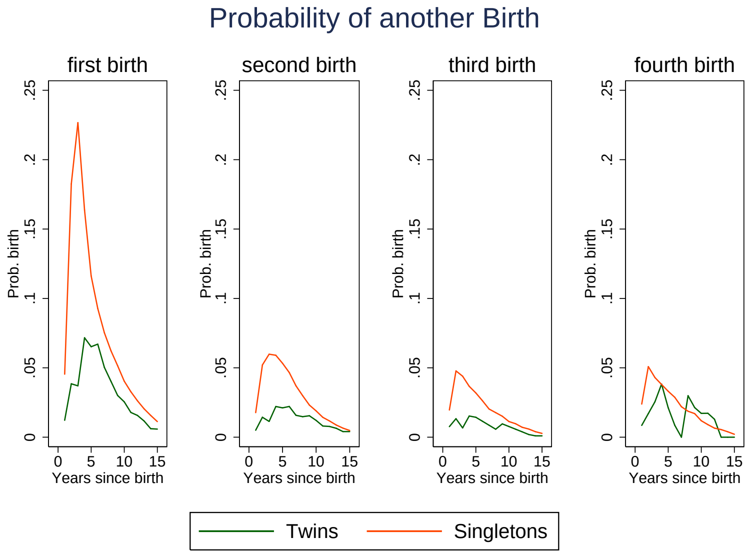

Figure 2: Probability of giving birth twin and nontwin mothers

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

Fig. 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 behaviour 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 behaviour on the lifepaths 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 behaviours. 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 carreers or reducing the amount of hours worked.

Figure 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 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 behaviour.
This is bound to upward bias our estimates of how twinning affects maternal long-term labour 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 behaviour 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 behaviour", 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 behaviour. 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 behaviour, 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 2 shows that, as expected, the difference in subsequent fertility behaviour 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 labour-market outcomes, this would be a very good indicator that our worries about substantial downward bias might be warranted.

4.2 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 [29]. The argumnent 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 include only first-birth mothers to include 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 labour market outcomes of mother i measured t years after her nth birth as Y_{itn}. We thus estimate the twins first approach as follows

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

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 this coefficient 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 behaviour, 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 are they able to account for subsequent fertility behaviour. Indeed in some of the previous papers estimating the long-term effects of twinning on maternal labour market outcomes, the authors do at times find positive effects of twinning on long-term labour market outcomes of mothers [24]. While this is not theoretically impossible, it is still rather unexpected. Taking into account the fact that the strongly differing subsequent fertility behaviour 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.

4.3 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 *n*th 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 labour-market outcomes. And in the first stage we instrument the number of children a mother had with a variable capturing whether she twinned at *n*th 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 labour market outcomes t years after birth we thus restrict the sample to all the mothers who had their *n*th 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 ran thus looks as follows

$$\hat{C}_{itn} = \gamma_{0tn} + \gamma_{1tn}Age_{itn} + \gamma_{2tn}Age^2_{itn} + \gamma_{3tn}Twins_{ni} + \mu_{tni} \tag{3}$$

The second stage regression then looks similar to 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 *n*th parity has on the number of children a person has at time t conditional on age. Angrist [4] has shown that in a model without covariates the Wald Estimate of the instrumented variable can be interpreted as follows:

$$\beta_{Wtn} = \frac{E[Y_{itn}|Twins_{tn} = 1] - E[Y_{itn}|Twins_{ni} = 0]}{E[C_{itn}|Twins_{tn} = 1] - E[C_{itn}|Twins_{ni} = 0]} \tag{4}$$

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 *n*th birth and mothers that gave birth to singletons at their *n*th birth t years after that *n*th birth occured. In the IV model, including 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. 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 their wages and employment. Since these effects are highly dependent 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 [23] 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 [3] 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 [23] 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 literature acknowledges three factors that might lead to correlation with the error term in the context of twinning regression. These are the effects of age, race and of in-vitro fertilization on twinning probabilities. The correlation with the error term of course stems from the fact that these three factors influence not only twinning probabilities, but at the same time, through a variety of channels 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 difference in subsequent fertility behaviour between twinning and non-twinning mothers is another factor which is bound to result in correlation of the instrument with the error term. These differences, ore 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. 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 labour-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.

5 Results

5.1 Main Results

We ran both the twins-first model described by 2 as well as the IV model described by 3 and 4 to see the results that the different methods common in the literature would deliver. We ran our estimations separately for four subsamples of mothers for each of the four birth parities n we consider. 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 3 and 4.

We thus went on to run separate regressions for each of the 2 outcome variables y, labour 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 [6].

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

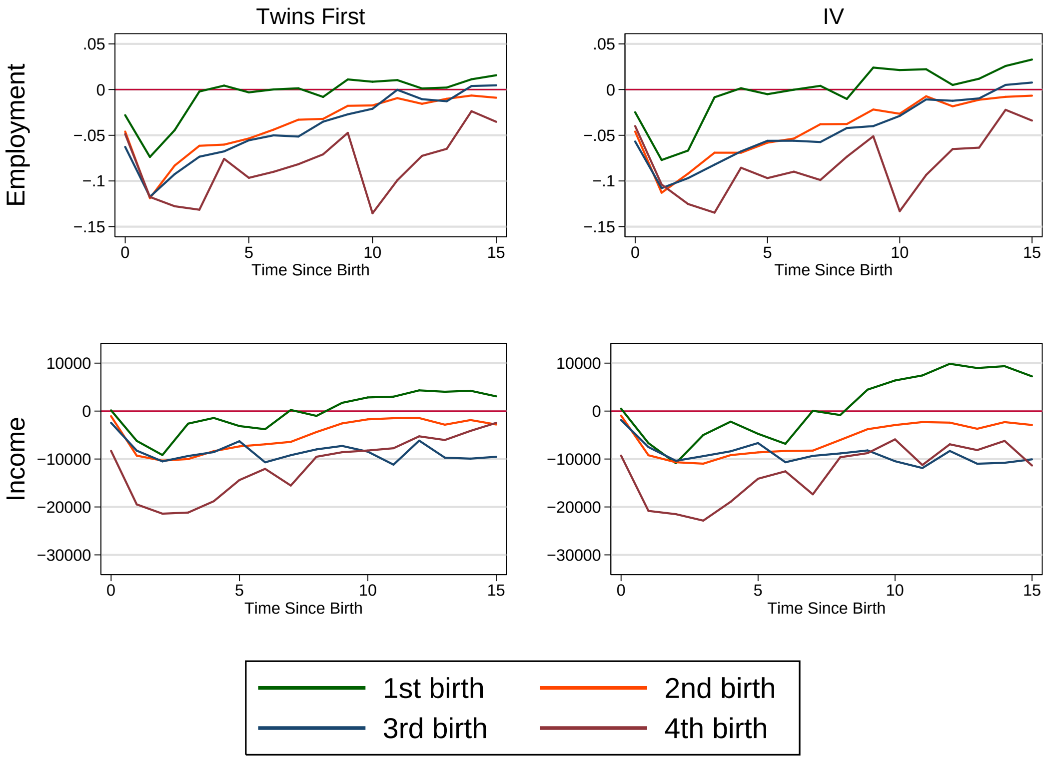

Figure 3: Development of Employment for twinnning and non-twinning mothers for selected cohorts

In the case of our twins-first model the parameter we are interested in is β_{3tn} from 2 and in the case of the of our IV regression it is β_{Wtn} from 4. The results of the 296 regressions are shown in Fig. 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 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. This is very much in line with what we predicted in the case that bias due to subsequent fertility behaviour 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 behaviour 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 behaviour 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 behaviour at these birth parities does not differ that strongly cannot entirely be explained by simply referring to subsequent fertility behaviour. The cautious conclusion at this point would be that the circumstantial evidence supports the theory that estimates at firstbirth are downward biased , but some of the differences we observe for the different parities might also be due to factors of selection and of changed household economics and decision making, that come with a higher number of children. To shed further light on how plausible it is that the differences are indeed driven by the biases introduced by subsequent fertility behaviour, 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 behaviour very similar to that of second and thirdbirth mothers and we also find similar effects on labour 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: 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

Table 4: Cumulative Effects, since birth, as estimated by IV

In order to gain a better grasp of the magnitude of the effects that an extra child has on labour-market outcomes, table 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 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.

Since including the 296 regressions from which the coefficients that are graphed in Fig. 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

Table 5: Regressions for Maternal Employment and Income

	1st birth			2nd birth			3rd birth			4th birth		
	0-5 yrs	6-10 yrs	11-15 yrs	0-5 yrs	6-10 yrs	11-15 yrs	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*	-0.0862***	-0.0497***	-0.00817	-0.104***	-0.0866***	-0.0862***
	(-4.45)	(0.28)	(1.37)	(-10.67)	(-5.20)	(-1.70)	(-6.52)	(-3.42)	(-0.57)	(-3.58)	(-2.64)	(-2.62)
age	0.157***	0.164***	0.166***	0.169***	0.176***	0.177***	0.169***	0.185***	0.200***	0.136***	0.158***	0.174***
	(120.74)	(71.44)	(92.60)	(113.65)	(93.17)	(63.72)	(59.02)	(47.09)	(34.71)	(21.54)	(18.38)	(12.96)
age2	-0.00229***	-0.00211***	-0.00193***	-0.00235***	-0.00218***	-0.00200***	-0.00216***	-0.00216***	-0.00215***	-0.00160***	-0.00176***	-0.00182***
	(-103.94)	(-56.40)	(-69.42)	(-98.71)	(-87.95)	(-65.46)	(-50.51)	(-44.62)	(-35.13)	(-17.58)	(-17.28)	(-13.31)
c	-1.892***	-2.311***	-2.716***	-2.113***	-2.546***	-3.003***	-2.314***	-2.935***	-3.782***	-1.856***	-2.472***	-3.099***
	(-114.07)	(-96.25)	(-60.73)	(-75.49)	(-51.89)	(-37.93)	(-36.06)	(-25.88)	(-21.72)	(-11.47)	(-8.52)	(-6.84)
Income												
kids	-4464.9**	-1639.1	9616.7**	-7969.9***	-6419.2***	-1897.8	-8046.3***	-9450.2***	-10551.0***	-18462.2***	-13399.9**	-6534.6
	(-2.35)	(-0.55)	(2.35)	(-6.28)	(-3.31)	(-0.81)	(-3.60)	(-3.13)	(-3.14)	(-4.15)	(-2.25)	(-0.84)
age	23451.8***	20330.4***	25203.2***	18350.4***	19957.3***	28586.4***	15239.4***	18747.7***	25999.8***	13937.2***	17845.8***	24566.8***
	(70.46)	(36.81)	(47.13)	(56.61)	(42.94)	(34.39)	(26.33)	(21.88)	(17.61)	(7.20)	(9.11)	(7.17)
age2	-310.2***	-225.0***	-245.9***	-226.8***	-214.0***	-286.5***	-167.0***	-190.0***	-253.1***	-138.1***	-175.4***	-231.5***
	(-53.52)	(-24.68)	(-28.68)	(-42.16)	(-33.62)	(-30.55)	(-18.90)	(-17.31)	(-15.58)	(-4.66)	(-7.22)	(-6.42)
c	-246917.9***	-220200.7***	-372810.7***	-166176.5***	-215352.5***	-422696.9***	-143227.5***	-209966.0***	-371122.1***	-104379.8***	-195752.7***	-383703.4***
	(-67.45)	(-39.59)	(-28.51)	(-29.63)	(-18.58)	(-18.05)	(-12.33)	(-8.98)	(-8.87)	(-2.80)	(-3.37)	(-3.45)
N	1416337	1408830	1407581	1135243	1131371	1129316	380280	378763	377128	85306	84873	84181

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

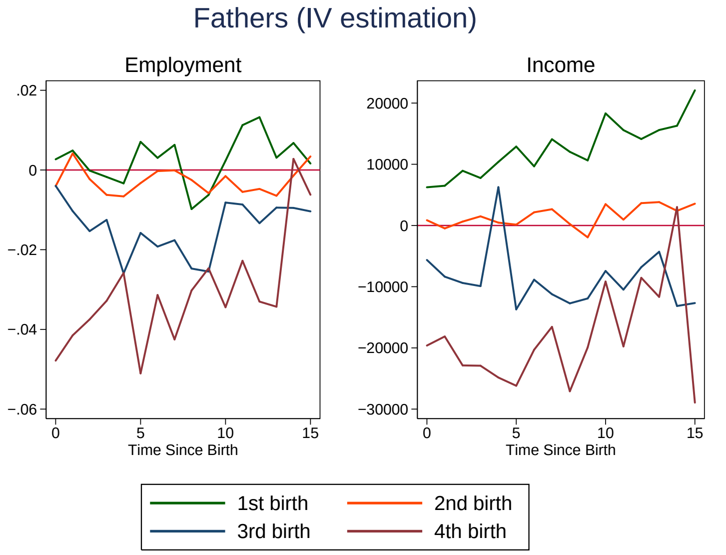

Figure 4: Instrumental Variable Estimates for G

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 6. The story that emerges from them is very similar to what we found when analyzing the graphs.

5.2 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.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 labour market outcomes. Angrist and Evans [5] 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 behaviour of fathers but where the labour-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 their parental labour-market outcomes, dependent on birth-order might change with birth-parity, for facts unrelated to subsequent fertility behaviour. Thus even though we have good reasons to believe that our maternal results are to an important extent driven by differences in subsequent fertility behaviour, 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 carreers 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 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 carreers 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

Figure 5: Effects of twinning by maternal income

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 5 shows the coefficients we obtained by estimating our IV model described by 3 and 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 [7] 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.

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

Table 6: Regressions for Maternal Employment and Income

	1st birth			2nd birth			3rd birth			4th birth		
	0-5 yrs	6-10 yrs	11-15 yrs	0-5 yrs	6-10 yrs	11-15 yrs	0-5 yrs	6-10 yrs	11-15 yrs	0-5 yrs	6-10 yrs	11-15 yrs
Employment												
kids	0.000937	0.00444	0.00782	-0.00528	-0.00318	-0.00254	-0.0109	-0.0201*	-0.00613	-0.0464**	-0.0472**	-0.0291
	(0.17)	(0.60)	(1.04)	(-1.10)	(-0.57)	(-0.43)	(-1.19)	(-1.85)	(-0.55)	(-2.15)	(-2.03)	(-1.16)
	(-4.45)	(0.28)	(1.37)	(-10.67)	(-5.20)	(-1.70)	(-6.52)	(-3.42)	(-0.57)	(-3.58)	(-2.64)	(-2.62)
age	0.0537***	0.0711***	0.0808***	0.0567***	0.0743***	0.0915***	0.0714***	0.0949***	0.121***	0.0917***	0.117***	0.142***
	(64.90)	(61.48)	(57.94)	(57.11)	(55.13)	(52.01)	(38.03)	(36.38)	(34.83)	(22.53)	(20.13)	(17.88)
age2	-0.00229***	-0.00211***	-0.00193***	-0.00235***	-0.00218***	-0.00200***	-0.00216***	-0.00216***	-0.00215***	-0.00160***	-0.00176***	-0.00182***
	(-103.94)	(-56.40)	(-69.42)	(-98.71)	(-87.95)	(-65.46)	(-50.51)	(-44.62)	(-35.13)	(-17.58)	(-17.28)	(-13.31)
c	-1.892***	-2.311***	-2.716***	-2.113***	-2.546***	-3.003***	-2.314***	-2.935***	-3.782***	-1.856***	-2.472***	-3.099***
	(-114.07)	(-96.25)	(-60.73)	(-75.49)	(-51.89)	(-37.93)	(-36.06)	(-25.88)	(-21.72)	(-11.47)	(-8.52)	(-6.84)
Income												
kids	-4464.9**	-1639.1	9616.7**	-7969.9***	-6419.2***	-1897.8	-8046.3***	-9450.2***	-10551.0***	-18462.2***	-13399.9**	-6534.6
	(-2.35)	(-0.55)	(2.35)	(-6.28)	(-3.31)	(-0.81)	(-3.60)	(-3.13)	(-3.14)	(-4.15)	(-2.25)	(-0.84)
age	23451.8***	20330.4***	25203.2***	18350.4***	19957.3***	28586.4***	15239.4***	18747.7***	25999.8***	13937.2***	17845.8***	24566.8***
	(70.46)	(36.81)	(47.13)	(56.61)	(42.94)	(34.39)	(26.33)	(21.88)	(17.61)	(7.20)	(9.11)	(7.17)
age2	-310.2***	-225.0***	-245.9***	-226.8***	-214.0***	-286.5***	-167.0***	-190.0***	-253.1***	-138.1***	-175.4***	-231.5***
	(-53.52)	(-24.68)	(-28.68)	(-42.16)	(-33.62)	(-30.55)	(-18.90)	(-17.31)	(-15.58)	(-4.66)	(-7.22)	(-6.42)
c	-246917.9***	-220200.7***	-372810.7***	-166176.5***	-215352.5***	-422696.9***	-143227.5***	-209966.0***	-371122.1***	-104379.8***	-195752.7***	-383703.4***
	(-67.45)	(-39.59)	(-28.51)	(-29.63)	(-18.58)	(-18.05)	(-12.33)	(-8.98)	(-8.87)	(-2.80)	(-3.37)	(-3.45)
N	1416337	1408830	1407581	1135243	1131371	1129316	380280	378763	377128	85306	84873	84181

t statistics in parentheses

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

5.3 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. We tested the exten 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 justtified when using more recent data. As we can see in table 7 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 7: 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 mght 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 are robust to specifications assuring that there is no endogeneity induced via IVF. We thus re-estimated our IV models using only births that happened between 1980 and 1986 when the role of in-vitro-fertilization, was very minor, or basically non-existent. Fig. 6 shows our results from those models for our maternal labour market variables.. What we find is very reassuring. The same picture, of increased negative effects of an extra birth on labour 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

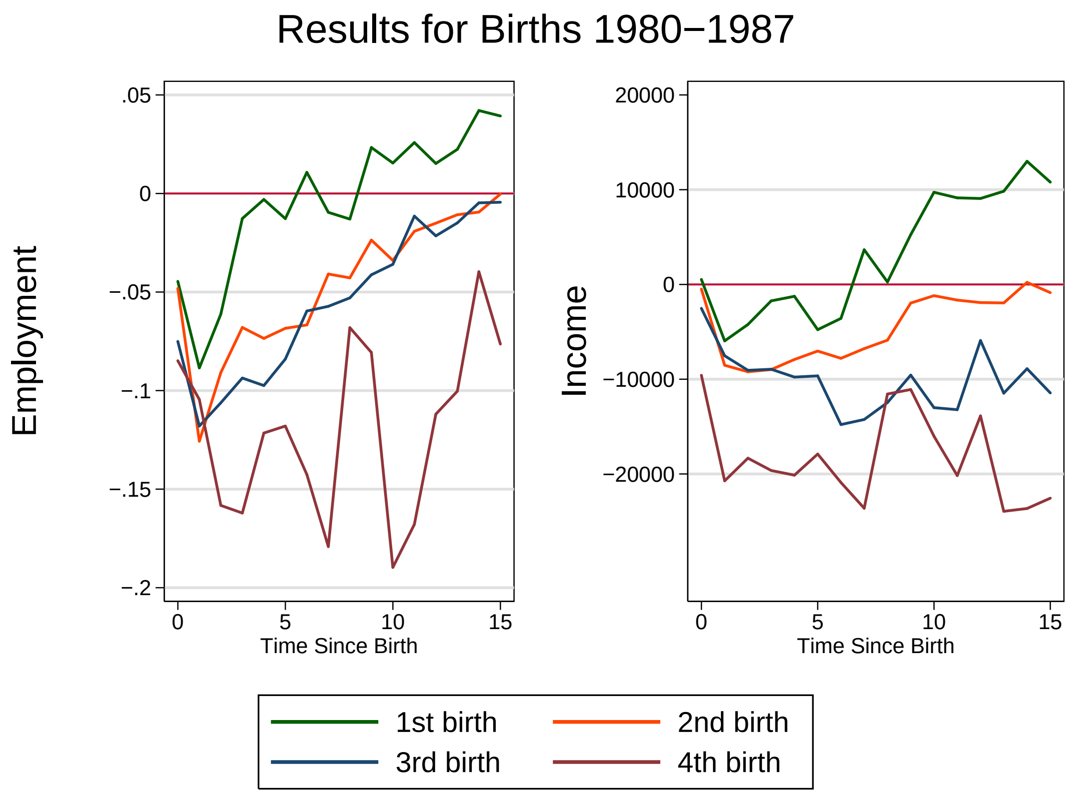

Figure 6: Instrumental Variable Estimates for Birth Cohorts 1980 to 1986

to conclude that our estimates were consistent with a story of considerable bias when looking at long-term labour-market outcomes at firstbirth re-emerged. Our results on the development of paternal labour-market incomes as well as on the effects of twinning, dependent on the income distribution also proved robust to this type of specification [8]

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 behaviour is bound to affect our estimates and 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 behaviour. One could easily enough argue that learning by the mothers makes

[8] results available on request

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 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. 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 labour 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 thos we found for other firstbirth mothers. If the different effects on labour market outcomes that we found between our firstbirth and our second-birth sample were however driven by different subsequent fertility behaviour 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 samplesize in the >35 sample our estimates are relatively noisy as can be seen in the greater volatility of the graph. Nevertheless it is very clear that the effects of twinning on labour-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. 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 labour market outcomes of their mothers were downward biased.

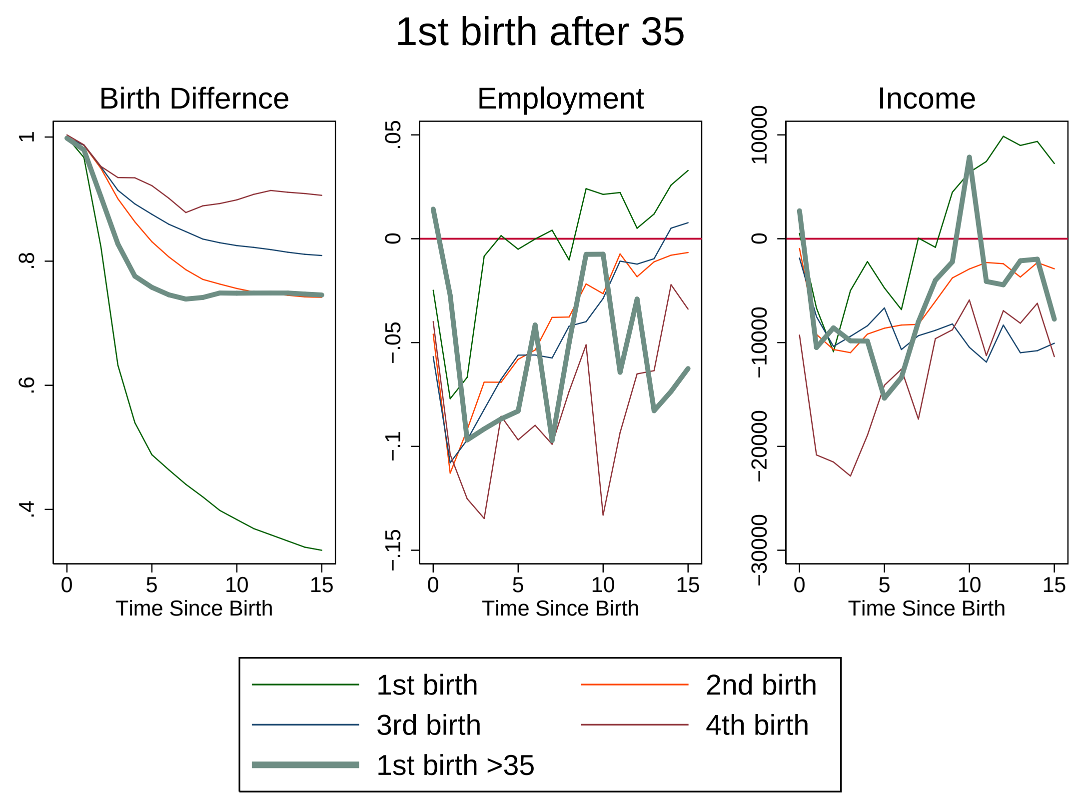

Figure 7: Comparing the effects of a firstbirth over the age 35 to full sample estimates for different birth parities

6 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 labour 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 behaviour 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 passes and subsequent fertility behaviour 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 behaviour 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 behaviour 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 labour-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 behaviour which we outlined. We also derive additional stylized facts on the effects of children on paternal labour 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 labour 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 labour 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.

References

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

[2] Joshua Angrist, Victor Lavy, and Analia Schlosser. Multiple experiments for the causal link between the quantity and quality of children. *Journal of Labor Economics*, 28(4):773–824, October 2010.

[3] Joshua Angrist, Victor Lavy, and Analia Schlosser. Multiple experiments for the causal link between the quantity and quality of children. *Journal of Labor Economics*, 28(4):773–824, October 2010.

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

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

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

[7] Gary S Becker and Nigel Tomes. Child endowments and the quantity and quality of children. *Journal of Political Economy*, 84(4):S143–62, August 1976.

[8] S Black, Paul Devereux, and Kjell Salvanes. 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, June 2005.

[9] Martin Browning. Children and household economic behavior. *Journal of Economic Literature*, 30(3):1434–75, September 1992.

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

[11] R. Carrasco. Binary choice with binary endogenous regressors in panel data: Estimating the effect of fertility on female labour participation. Technical report, 1998.

[12] Julio Cáceres-Delpiano and Marianne Simonsen. The toll of fertility on mothers wellbeing. *Journal of Health Economics*, 31(5):752–766, 2012.

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

[14] Anna Christina D'Addio and Marco Mira d'Ercole. Trends and determinants of fertility rates: The role of policies. OECD Social, Employment and Migration Working Papers 27, OECD Publishing, September 2005.

[15] Esping-Andersen and Gosta. 2. A Child-Centred Social Investment Strategy. *Why We Need a New Welfare State*, pages 26–68.

[16] Gosta Esping-Andersen, Irwin Garfinkel, Wen-Jui Han, Katherine Magnuson, Sander Wagner, and Jane Waldfogel. Child care and school performance in denmark and the united states. *Children and Youth Services Review*, 34(3):576 – 589, 2012. <ce:title>Comparative Child and Family Policy</ce:title>.

[17] James Feyrer, Bruce Sacerdote, and Ariel Dora Stern. Will the stork return to europe and japan? understanding fertility within developed nations. *The Journal of Economic Perspectives*, 22(3), 2008.

[18] Claudia Goldin. Career and family: College women look to the past. Working Paper 5188, National Bureau of Economic Research, July 1995.

[19] Jeff Grogger and Stephen G. Bronars. 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, June 2001.

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

[21] Jill Hardin, Suzan L. Carmichael, Steve Selvin, and Gary M. Shaw. Trends in the probability of twins and males in california, 1983 to 2003. *Twin Research and Human Genetics*, 12:93–102, 1 2009.

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

[23] Guido W. Imbens and Joshua D. Angrist. Identification and estimation of local average treatment effects. *Econometrica*, 62(2):pp. 467–475, 1994.

[24] Joyce P. Jacobsen, James Wishart Pearce III, and Joshua L. Rosenbloom. 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, 1999.

[25] Florence Jaumotte. Female labour force participation: Past trends and main determinants in oecd countries. OECD Economics Department Working Papers 376, OECD Publishing, December 2003.

[26] Thomas Liebig. The labour market integration of immigrants in denmark. OECD Social, Employment and Migration Working Papers 50, OECD Publishing, 2007.

[27] J. Mincer. *Opportunity costs and income effects*. Stanford University Press, Stanford, CA, 1963.

[28] OECD. Oecd family database. 2012.

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

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

[31] James P. Smith and Michael P. Ward. Time-series growth in the female labor force. *Journal of Labor Economics*, 3(1):pp. S59–S90, 1985.

[32] Jeroen SMITS and Christiaan MONDEN. Twinning across the developing world. *PLoS ONE*, 6(9), 2011.

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

[34] Tine Westergaard, Jan Wohlfahrt, Peter Aaby, and Mads Melbye. Population based study of rates of multiple pregnancies in denmark, 1980-94. *BMJ*, 314(7083):775, 3 1997.