IFAU
Institute for Evaluation of Labour
Market and Education Policy

Increasing the credibility of the twin birth instrument

Helmut Farbmacher
Raphael Guber
Johan Vikström

WORKING PAPER 2016:10

The Institute for Evaluation of Labour Market and Education Policy (IFAU) is a research institute under the Swedish Ministry of Employment, situated in Uppsala. IFAU's objective is to promote, support and carry out scientific evaluations. The assignment includes: the effects of labour market and educational policies, studies of the functioning of the labour market and the labour market effects of social insurance policies. IFAU shall also disseminate its results so that they become accessible to different interested parties in Sweden and abroad.

IFAU also provides funding for research projects within its areas of interest. The deadline for applications is October 1 each year. Since the researchers at IFAU are mainly economists, researchers from other disciplines are encouraged to apply for funding.

IFAU is run by a Director-General. The institute has a scientific council, consisting of a chairman, the Director-General and five other members. Among other things, the scientific council proposes a decision for the allocation of research grants. A reference group including representatives for employer organizations and trade unions, as well as the ministries and authorities concerned is also connected to the institute.

Postal address: P.O. Box 513, 751 20 Uppsala
Visiting address: Kyrkogårdsgatan 6, Uppsala
Phone: +46 18 471 70 70
Fax: +46 18 471 70 71
ifau@ifau.uu.se
www.ifau.se

Papers published in the Working Paper Series should, according to the IFAU policy, have been discussed at seminars held at IFAU and at least one other academic forum, and have been read by one external and one internal referee. They need not, however, have undergone the standard scrutiny for publication in a scientific journal. The purpose of the Working Paper Series is to provide a factual basis for public policy and the public policy discussion.

ISSN 1651-1166

Increasing the credibility of the twin birth instrument[a]

by

Helmut Farbmacher[b], Raphael Guber[c] and Johan Vikström[d]

June 9, 2016

Abstract

Twin births are an important instrumental variable for the endogenous fertility decision. However, in many economic settings, twins are not exogenous as dizygotic twinning is known to be correlated with maternal characteristics and fertility treatments. Following the medical literature, we assume that monozygotic twins are exogenous, and construct a new instrument, which corrects for the selection bias although monozygotic twinning is usually unobserved. We use longitudinal administrative data from Sweden and US census data and show that the usual twin instrument is not only related to observed but also to unobserved determinants of economic outcomes, while our new instrument is not. We demonstrate the relevance of our new instrument in two labor market applications and find that the classical twin instrument underestimates the true negative effect of fertility on labor force participation and earnings.

JEL-codes: C26, J13, J22

[a]We are grateful for helpful suggestions from Marcus Eliason, Martin Nybom and Martin Weidner and participants at the International Association for Applied Econometrics conference, the Econometric Society World congress, the Royal Economic Society annual conference, the Verein fuer Socialpolitik annual conference and IFAU. Guber acknowledges funding through the International Doctoral Program "Evidence-Based Economics" of the Elite Network of Bavaria. Vikström acknowledges support from the Ragnar Söderberg foundation.

[b]Munich Center for the Economics of Aging, Max Planck Society

[c]Munich Center for the Economics of Aging, Max Planck Society

[d]IFAU and UCLS, E-mail: johan.vikstrom@ifau.uu.se

Table of contents

1 Introduction

As fertility decisions are endogenous, most papers on how family size affects maternal and child outcomes use instrumental variable techniques. One commonly employed instrument for the number of children is twin births. Early studies that use the twin instrument to study maternal outcomes include Rosenzweig & Wolpin (1980a); Bronars & Grogger (1994); Angrist & Evans (1998); Jacobsen et al (1999).[1] The twin instrument has also been used to study the prediction of the Becker & Lewis (1973) quantity–quality model, that family size has a negative effect on children's economic outcomes (Rosenzweig & Wolpin, 1980b; Black et al, 2005; Cáceres-Delpiano, 2006; Angrist et al, 2010). However, it has been questioned if having twins—particularly dizygotic twins—really is a random event. In particular, it has been shown that dizygotic twinning depends on, for example, maternal age, height, weight, race, and the use of fertility treatments (Reddy et al, 2005; Fauser et al, 2005). On the other hand, monozygotic (identical) twin births are considered a random event (Tong & Short, 1998; MacGillivray et al, 1988), since they are the result of the random and spontaneous division of a single fertilized egg (e.g., Hall, 2003)[2].

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

Initially we use longitudinal data from Sweden to show that twin births are correlated with observed and unobserved maternal characteristics and that this correlation is stronger in more recent cohorts. To analyse the selection on unobservables, we use information

[1] More recent studies include, for instance, Vere (2011) on parental labor supply and Cáceres-Delpiano & Simonsen (2012) on its effects on general well-being.

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

about pre-pregnancy labor force participation, labor income, and hospitalizations, and conclude that these pre-pregnancy outcomes predict future twin births. This selection is likely to be even more pronounced in data from the US, where twin rates are almost twice as high as those in Sweden. Using US census data, Angrist & Evans (1998) show that twins are more common among older and highly educated women and among certain races. This is also supported by other studies in both economics and medicine (see e.g., Figlio et al, 2014, Black et al, 2005 or Martin & Park, 1999). We emphasize, however, that these concerns only apply to dizygotic twin births and not to monozygotic twin births.

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

We also discuss ways to relax our main assumption using instead that monozygotic twinning is less endogenous than dizygotic twinning. Here, we add to the concept of imperfect instruments by Nevo & Rosen (2012) and Conley et al (2012), by considering misclassified discrete instrumental variables. Nevo & Rosen (2012) examine identification under different assumptions, for instance, that the correlation between the instrument and the error term is less than the correlation between the endogenous variable and the error term. Conley et al (2012) consider identification and inference for different strategies that use prior information about how close the exclusion restriction is to being satisfied.

There are several reasons why our contribution is important. Firstly, twin births provide an unexpected fertility shock and are also a particularly relevant instrument, as, in many cases, it shifts the number of children by one.[3] Secondly, as already mentioned, the twin instrument has been used in several settings, including studies on fertility and maternal outcomes and studies of the child quality-quantity hypothesis.

Thirdly, since the mid-1970's we have seen a rise in the twinning rate, caused by delayed childbearing and an increasing need for fertility treatments (Martin et al, 2012; Fauser et al, 2005). Since the decision to undergo fertility treatment is an endogenous choice, which is clearly affected by the wish or need to postpone motherhood, it is even more likely that the twin births induced by in-vitro fertilization (IVF) are correlated with important socioeconomic characteristics. For instance, Braakmann & Wildman (2014) show that instrumental variables estimates with and without information on fertility treatments might differ substantially in applications to female labor supply and the child quantity–quality relation.[4] This suggests that mothers with twins have become an increasingly selective sample, which poses a threat to the identification of causal effects using the classical twin instrument.

Fourthly, there are only a few other potential variables which can serve as an instrument for endogenous fertility decisions. A commonly used instrument is parental preference for a mixed sex composition of children, but this instrument has been criticized because it uses a planned (as opposed to an unplanned) change in fertility to identify the causal effect (Butcher & Case, 1994; Rosenzweig & Wolpin, 2000). It identifies the local treatment effect for parents that actually have preferences for a mixed sex offspring, and Agüero & Marks (2008) note that these women may differ systematically from the population at large. Other previously used instruments for fertility are natural infertility (Agüero & Marks, 2008), successful IVF treatment (Lundborg et al, 2014), and, in cultures with strong son preferences, the sex of the first child (Lee, 2008).

Besides the non-randomness of twin births another concern with the twin instrument, raised by Rozenzweig and Zhang (2009), is that twins have inferior endowments at birth,

[3] Another advantage is that, by the instrument's nature, the monotonicity assumption is fulfilled, and, as there can be no defiers or never-takers, the twin IV estimate can be interpreted as the average treatment effect on the untreated when the endogenous variable is a dummy for one more child (Angrist et al, 2010).

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

such as lower APGAR scores and lower birth weight, than singletons. If these differences induce parents to reallocate resources across their children this will violate the exclusion restriction in studies that uses the twin instrument to study quantity-quality effects on non-twin siblings. Rozenzweig and Zhang (2009) find that such differential birth endowment effects are important, while Angrist et al. (2010) find no evidence that would invalidate the exclusion restriction. Another concern with the twin instrument is that the close spacing of twins makes their child-rearing more equal, leading to economics of scale in the child quality production. These concerns are mainly relevant for studies using the twin instrument to study the child quality-quantity hypothesis and are arguably less of a concern for studies that use the twin instrument to study fertility effects on maternal outcomes. Finally, note that the assumption that monozygotic twins are at least less endogenous than dizygotic twins is still valid if the birth endowment effect and the economics of scale effect are the same for monozygotic and dizygotic twins.

We use both Swedish and US data to illustrate our new approach. We revisit the study by Angrist & Evans (1998) and use their data on mothers from the 1980 US census. One result is that both the classical twin instrument and the same-sex twin instrument underestimate the true negative effect of fertility on labor earnings. This confirms that dizygotic twin mothers are a positively selected sample, partly because high earners are more likely to delay childbearing and hence have a higher risk to get twins. We obtain similar results using Swedish register data both for mothers who got their first child before the strong rise in fertility treatments (1987–1990) and for mothers who got their first child during a later period with substantially higher twin rates (2000–2003).

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

Table 1: Summary statistics for our sample of Swedish mothers

	1987–1990		2000–2003	
	Mean	SD	Mean	SD
# mothers	175,011		155,392	
Socioeconomic characteristics				
Age (at first birth)	26.278	4.476	28.772	4.750
Less than nine years of schooling	0.0043	0.0654	0.0043	0.0654
Nine years of schooling	0.2130	0.4094	0.1201	0.3251
Two year high school	0.4474	0.4972	0.1608	0.3673
Three year high school	0.1336	0.3402	0.3108	0.4628
University or college $<$ 3 years	0.1250	0.3307	0.1706	0.3762
University or college $\geq$ 3 years	0.0753	0.2639	0.2294	0.4204
Phd education	0.0012	0.0346	0.0038	0.0615
Pre-pregnancy outcomes (two years before first birth)				
Labor force participation	0.9726	0.1633	0.9151	0.2787
Log labor income	11.450	0.8724	11.634	1.139
Hospitalization	0.1183	0.4492	0.0772	0.3711
Twin indicators				
Twins ($\ddot{z}_t$)	0.0103	0.1010	0.0188	0.1358
Same-sex Twins ($\dot{z}_t$)	0.0073	0.0853	0.0117	0.1077

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

2 Zygosity and selection on (un)observables

2.1 Data

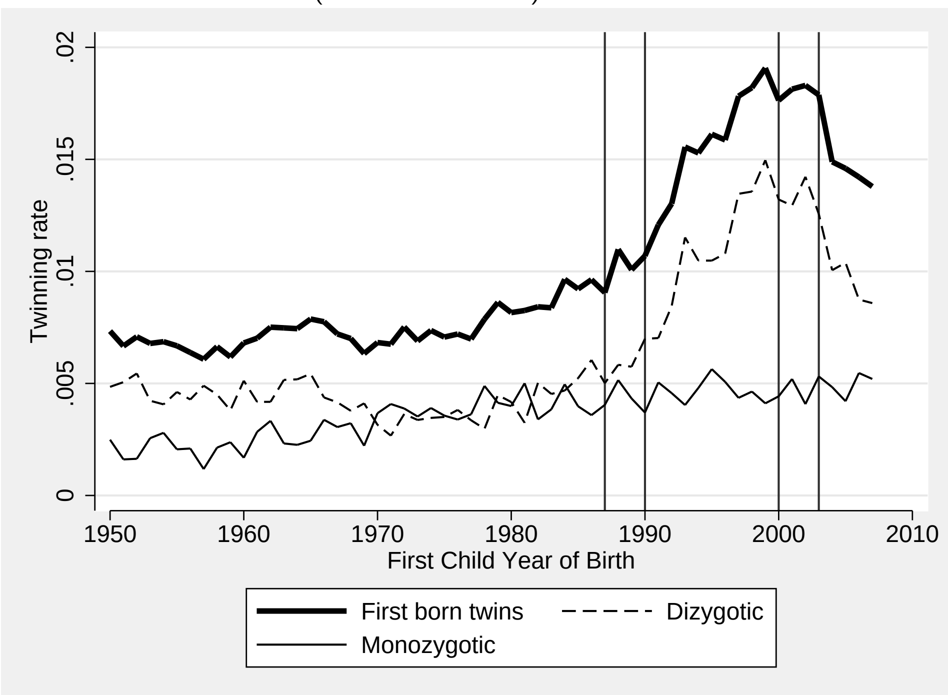
We use Swedish register data (Ekbom, 2011) to assess the importance of selection on observable and unobservable variables. The multi-generational register links individuals to their biological mother and father and contains information on the year and month of birth, which we use to construct information on twin births. The population register contains yearly information on labor income, labor force participation and education. The Swedish National Patient Register provides information on all episodes of in-patient care in Sweden between 1987 and 2005. Our sample comprises all mothers who got their first child either in the years 1987 to 1990 or 2000 to 2003, which gives us roughly 45,000 women per year.[5] *Table 1* gives some descriptive statistics of our data set. For instance, the maternal age at first birth was around 26.3 in the early period and 28.8 around 2001, reflecting the well-documented delay in childbearing.

[5]Because of data limitations, 1987 is the earliest year we can use for our analysis.

2.2 Twin births in Sweden and the US

To investigate the changes of twinning in Sweden over time, *Figure 1* shows the twin rates across the first child's year of birth.[6] The overall twinning rate remains fairly constant between 1950 and 1980 but increases thereafter. While the rather mild rise between 1980 and 1990 can be attributed to delayed child bearing, the steep increase since 1990 follows the availability of IVF. The drop in 2003 follows a recommendation by the Swedish National Board of Health and Welfare regarding the method of elective single embryo transfers (SET), which proceeds by implanting one fertilized egg at a time, instead of several eggs at once, as was done before (Bergh, 2005).

Figure 1: Twin rate in Sweden (firstborn children)



As can be seen in *Figure 1*, the earlier time period (i.e., 1987–1990) we are investigating was just at the beginning of a strong rise in overall twin births (thick solid line). IVF was rather unusual at this time. The later time period (i.e., 2000–2003), however, is associated with substantially higher twin rates, which are mainly caused by increased fertility treatments. In particular, from 1990 to 2000, the rate of dizygotic twins almost

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

tripled while the monozygotic rate remained fairly constant.

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

2.3 Selection on observable and unobservable characteristics

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

While we can easily determine whether there is any selection on observable characteristics, testing for selection on unobservables is by definition impossible. However, as many economic determinants are inherently persistent, we can assess the importance of the selection on unobservables by using pre-pregnancy outcomes. That is, we can test whether, conditionally on observable characteristics, twin mothers and non-twin mothers were already different before their first pregnancy. Our pre-pregnancy outcomes are labor force participation, yearly labor income, and hospitalizations two years before the first birth.[7] At this point in time, the future mothers have no children. They might not even know that they will have kids in two years, and they surely do not know that they will have

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

twins. Therefore the pre-pregnancy outcomes should be causally unaffected by the twin births and the only reason for a pre-pregnancy difference between the twin and non-twin groups is selection on unobserved characteristics.

Table 2 shows how our observed socio-economic characteristics and the pre-pregnancy outcomes correlate with twin births. Initially we regress twin indicators on mother's age at first birth and level of education, and report the overall F-statistic of joint significance. From column 1 we see that the usual twin indicator, $\ddot{z}$, correlates with these observables in both time periods, indicating that twin births are not a random event. The F-statistic increases strongly from 16.29 in 1987–1990 to 51.84 in 2000–2003. This reflects the strong rise in twin rates and the increased selection because of fertility treatments and delayed childbearing among mothers with high career preferences. Interestingly, when we use the improved same-sex twin indicator, $\dot{z}$ (see column 2), the value of the overall F-test statistic decreases by more than half in both periods (from 16.29 to 7.43 and from 51.84 to 20.75). This indicator variable excludes all opposite-sex twins which cannot be monozygotic.[8,9] Thus, when we exclude twins who have to be dizygotic and thereby implicitly increase the fraction of monozygotic twins, we see a lower dependence of the instrument on socio-economic characteristics.

We obtain similar patterns for the pre-pregnancy outcomes. *Table 2* shows that—conditionally on the set of covariates—there still are significant differences between women with and without future twins.[10] In the years 2000 to 2003 (lower panel), women were significantly more likely to be employed and had higher incomes two years before the birth of their twins. The probability of being hospitalized was increased in both time

[8] Note that we did not drop the opposite-sex twins from our analysis but instead used the more precise measure of monozygotic twinning. One could in principle also think about dropping the opposite-sex twins but the results should be almost similar due to the low frequency of twinning compared to singleton births. This is the case for our results, which are available from the authors upon request.

[9] Note that the F-statistic would decline anyway when the fraction of twins declines, even if we were to randomly exclude some of the twins. To further investigate this relation, we randomly exclude the fraction of opposite-sex twins in a simulation. Using 500 replications, we see an average drop of the F-test statistic to 11.71 around 1990 and to 32.73 around 2000. This is still distinctly larger than the F-test statistics of 7.43 and 20.75 which we obtain from the regressions on the same-sex twins.

[10] Note that each estimate comes from a single linear regression with the different twin instruments as the explanatory variable and controlling for the observed covariables.

Table 2: Assessing the importance of selection on observables and unobservables

	Twins	Same-sex twins	New instruments	
	$\ddot{z}_t$	$\dot{z}_t$	$\bar{z}_t(0)$	$\bar{z}_t(\theta_{min})$
			1987–1990	
Level of education and age				
F-statistic	16.29***	7.43***	1.06	0.85
Pre-pregnancy outcomes (two years before first birth)				
Labor force participation	-0.0018	-0.0020	-0.0010	-0.0007
	(0.0040)	(0.0047)	(0.0040)	(0.0037)
Log labor income	-0.0099	-0.0159	-0.0127	-0.0111
	(0.0179)	(0.0220)	(0.0177)	(0.0161)
Hospitalization	0.0400**	0.0185	-0.0140	-0.0180
	(0.0155)	(0.0178)	(0.0155)	(0.0144)
			2000–2003	
Level of education and age				
F-statistic	51.84***	20.75***	1.04	1.03
Pre-pregnancy outcomes (two years before first birth)				
Labor force participation	0.0111**	0.0118**	0.0037	0.0035
	(0.0046)	(0.0058)	(0.0045)	(0.0045)
Log labor income	0.0582***	0.0505**	0.0055	0.0047
	(0.0174)	(0.0213)	(0.0172)	(0.0170)
Hospitalization	0.0142**	0.0125	0.0015	0.0013
	(0.0063)	(0.0079)	(0.0062)	(0.0061)

Notes: For the pre-pregnancy, each cell reports estimates from a separate regression. LPM estimates using the sample of mothers described in the data section. All models also include year fixed effects. Level of education (7 categories) and age as a set of indicators. $\theta_{min} = 0.21$ in 1987–1990 and $\theta_{min} = 0.04$ in 2000–2003. Robust standard errors in parentheses. ***,**,* indicate significance at the 1%, 5%, and 10% level, respectively. Labor income is in SEK. Hospitalization is an indicator for at least one in-patient care episode. Number of observations for LFP: 175,011 (1987–1990) and 155,392 (2000–2003). Number of observations for ln(inc): 170,215 (1987–1990) and 142,201. Number of observations for hospitalization: 91,924 (1987–1990) and 155,392 (2000–2003).

periods.[11] The significant twin coefficients suggest that there are other (potentially persistent) unobservable variables which may confound estimates based on the conventional definition of the twin instrument. Similar to the results for selection on observables, these differences become less significant when we use the improved same-sex twin indicator. We will now turn to our methodological contribution and here we will also discuss the results in columns 3 and 4 of *Table 2*.

[11] Since hospitalization is known between 1987 and 2005, we can include only mothers that gave birth in 1989 or 1990 in the earlier time period sample.

3 Learning from monozygotic twins

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

$$y_i = \beta x_i + u_i, \qquad (1)$$

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

Let z^* be an indicator pointing to all mothers with monozygotic twins at their first birth.[13] Monozygotic twinning is most often unobservable—even for the parents. In one-third of identical twins, each fetus has its own placenta, which is also the case for all dizygotic twins (Bomsel-Helmreich & Al Mufti, 2005). Without further tests, these identical twins cannot be distinguished from fraternal twins unless they have opposite sexes. Typically, neither administrative data sets (like census data) nor surveys contain information on monozygosity. Therefore, we assume having data only on the classical twin indicator, which we denote by $\ddot{z}$, indicating both monozygotic and dizygotic twin births, and an indicator for the sex composition of the first two children (SX). Following

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

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

Angrist & Evans (1998), the latter variable is defined as $SX = s_1 s_2 + (1-s_1)(1-s_2)$, where s_1 and s_2 refer to male first-born and second-born children. Note that SX points to all siblings with the same sex, not only to like-sex twins. Since opposite sex twins can never be monozygotic, we can also define a more precise measure for monozygosity, namely the same-sex twin indicator $\dot{z}$. Define

$$\dot{z} = z^* + \dot{e},$$
$$\ddot{z} = z^* + e = z^* + \dot{e} + \ddot{e} = \dot{z} + \ddot{e},$$

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

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

A. 1 *monozygotic twinning is "less endogenous" than dizygotic twinning:*

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

twinning is relevant:

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

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

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

A. 2 *Weinberg (1901)'s rule:*

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

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

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

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

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

Assumption A.3 states that the same-sex instrument is exogenous within the group of mothers with dizygotic twins. The mean independence of the sex composition of siblings is generally considered as a more reliable exogeneity assumption compared to the mean independence of twinning (see the discussions in Angrist & Evans, 1998 and Black et al, 2010). In that sense, our approach combines the best of two instrumental variables: external validity from twinning and the more credible exogeneity assumption of the same sex instrument.

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

Lemma 1 *If d and w are random variables where d is binary and $E(w) = 0$, then*

$$\sigma_{wd} = \pi_d E(w|d = 1)$$

Proof of L. 1.

$$\begin{aligned}\sigma_{wd} &= E(wd) - E(w)E(d) = E(wd) \\ &= (1-\pi_d)E(wd|d=0) + \pi_d E(wd|d=1) \\ &= \pi_d E(w|d=1).\end{aligned}$$

where the second equality follows from $E(w) = 0$. The Law of Iterated Expectations gives the third equality and the conclusion follows from $E(wd|d = 0) = 0$ and $E(wd|d = 1) = E(w|d = 1)$. Q.E.D.

Using z^* as instrument, we asymptotically get

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

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

$$\begin{aligned}\beta_{\dot{z}}^{IV} &\equiv \frac{\sigma_{y\dot{z}}}{\sigma_{x\dot{z}}} \overset{A.1}{=} \beta + \frac{\pi_{z^*}\theta^* E(u|e=1)}{\sigma_{x\dot{z}}} + \frac{\sigma_{u\dot{e}}}{\sigma_{x\dot{z}}}, \\ \beta_{\ddot{z}}^{IV} &\equiv \frac{\sigma_{y\ddot{z}}}{\sigma_{x\ddot{z}}} \overset{A.1}{=} \beta + \frac{\pi_{z^*}\theta^* E(u|e=1)}{\sigma_{x\ddot{z}}} + \frac{\sigma_{u\dot{e}}}{\sigma_{x\ddot{z}}} + \frac{\sigma_{u\ddot{e}}}{\sigma_{x\ddot{z}}},\end{aligned} \tag{2}$$

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

Lemma 2 *If Assumption A.2 and A.3 hold, then*

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

Proof of L. 2. Note that by the definitions of $\dot{e} = SX \times e$ and $\ddot{e} = (1 - SX) \times e$, it follows that $E(u|\dot{e} = 1) = E(u|SX = 1, e = 1)$ and $E(u|\ddot{e} = 1) = E(u|SX = 0, e = 1)$. Furthermore,

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

Using Lemma 2 we can derive the following moment condition

$$E(u\bar{z}(\theta)) = 0, \tag{3}$$

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

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

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

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

Assuming that monozygotic twinning is at least less correlated with unobserved characteristics than dizygotic twinning (i.e., $-1 < \theta^* < 1$), we can obtain a set of estimates for β under different assumptions about the degree of endogeneity of monozygotic twinning. For this we construct $\bar{z}(\theta)$ for a grid of values of the endogeneity parameter θ (in between -1 and 1) and calculate the 2SLS estimate separately for each of these variables. This procedure is similar to the idea of imperfect instruments in Nevo & Rosen (2012). They argue that if z is less endogenous than x, the ratio of the correlations between z and u and between x and u must be between zero and one, i.e., $\lambda = \rho_{zu}/\rho_{xu} \in (0,1)$. Knowledge of λ would enable the construction of an exogenous instrument, but in its absence, one can use any reasonable value or a set of values between zero and one to construct new instruments and to bound the causal effect.

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

We now return to the results on the selection on observables in *Table 2*, columns 3 and 4 for both time periods, to assess how the new instrument correlates with mothers' observed characteristics. We use $\theta = 0$ in column three, while a grid search over $\theta \in (-1,1)$ reveals that the lowest overall F-statistic is obtained by setting $\theta_{min} = .21$ for the first time period (1987–1990) and $\theta_{min} = .04$ for the latter period (2000–2003). The results for θ_{min} are reported in column four. The F-statistic is approximately one for our proposed instruments, indicating that the observables cannot explain the variation in our new instruments. Turning to the remaining results in *Table 2*, we find that selection on unobservables is reduced as well, in particular in the 2000 to 2003 period. None of the proposed instruments is correlated with the pre-pregnancy outcomes. Moreover, θ_{min} decreases from .21 for the earlier time period to .04 for the later period, which is due to the relation $E(u|z^* = 1) = \theta^* E(u|e = 1)$. If monozygotic twinning is not influenced by socio-economic characteristics and fertility treatments (as *Figure 1* suggests), $E(u|z^* = 1)$ should be constant over time, which implies that θ has to decrease as $E(u|e = 1)$ increases. Thus, the decline in θ_{min} suggests that dizygotic twinning becomes more endogenous over time.

4 Empirical applications

4.1 Swedish register data

We now apply our new instrument to the two Swedish cohort-based samples introduced in Section 2. Our outcome variables are labor force participation and yearly labor income one year after the birth of the first child. We are interested in the effects of having more than one child one year after the first birth and use either the classical ($\ddot{z}$), the same-sex ($\dot{z}$) or our ($\bar{z}$) twin indicator as instruments.[14] If, as we have demonstrated in Section 2, twinning is more endogenous in the 2000–2003 than in the 1987–1990 cohort, we expect

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

the 2SLS coefficients obtained by using our new instrument to differ more markedly from those obtained by using the classical or the same-sex twin instrument in the former than in the latter sample. We control for mothers' age at first birth and education, as well as time (year) fixed effects.[15] Note that, within the 1987-1990 cohort, about 10% of the mothers had more than one child the year after the first birth, while this figure is 8% for the 2000–2003 cohort.

For both samples, our new instrument gives the strongest effect out of all IV regressions, correcting for positively selected mothers with dizygotic twins (*Table 3*).[16] The coefficients differ more strongly in the 2000–2003 cohort which contains a large fraction of dizygotic twin mothers. For example, for the earlier cohort, the estimated effect on labor force participation (upper panel) is -6.0% when using the classical twin instrument and -9.4% when using the new instrument—a relative difference of more than half. For the more recent cohort, the correction is even stronger with a relative difference by a factor of 2.7. This pattern is repeated for yearly labor income (lower panel) with a lower magnitude.

4.2 1980 US Census data

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

AE use the usual twin indicators $\ddot{z}$, $\dot{z}$, and an indicator for same-sex siblings. To this we add our two instruments. $\bar{z}(0)$ is constructed by assuming that $\theta = 0$, i.e., monozygotic twins are uncorrelated with the structural error term. In practice, this delivers an instrument which takes on a value of -1 for all opposite-sex twins, a value of 1 for same-sex

[15]Mothers education is taken from the year of their first birth. If this was missing, we use the information from up to seven subsequent years.

[16]Note that the first-stage F-statistic seems extremely large for the $\ddot{z}$ and $\dot{z}$ instruments, which comes from the fact that as we are looking at short run outcomes only one year after first birth, about 22% (10%) of all mothers that have more than child gave birth to twins in the 1987–1990 (2000–2003) cohort.

Table 3: Effect of having more than one child one year after birth - Swedish data

	OLS	2SLS		
		$\ddot{z}$	$\dot{z}$	$\bar{z}$
		Labor Force Participation		
First child born between 1987 and 1990 (N=184,587)				
More than one child	-0.074***	-0.060***	-0.070***	-0.094***
	(0.003)	(0.010)	(0.012)	(0.023)
First stage F-statistic		1,454,998	1,364,259	416
First child born between 2000 and 2003 (N=165,302)				
More than one child	-0.097***	-0.024***	-0.038***	-0.089***
	(0.004)	(0.008)	(0.010)	(0.031)
First stage F-statistic		1,909,464	1,677,854	221
		Log(Yearly Labor Income)		
First child born between 1987 and 1990 (N=158,827)				
More than one child	-0.453***	-0.461***	-0.471***	-0.495***
	(0.011)	(0.037)	(0.044)	(0.090)
First stage F-statistic		1,391,692	1,308,771	309
First child born between 2000 and 2003 (N=130,020)				
More than one child	-0.328***	-0.404***	-0.438***	-0.578***
	(0.015)	(0.031)	(0.040)	(0.133)
First stage F-statistic		1,753,934	1,493,759	142

Notes: Robust standard errors in parentheses. *** p<0.01. All coefficient and standard error pairs come from separate regressions. Outcomes and indicator for having more than one child are measured one year after birth of first child. $\ddot{z}$ is an indicator equal to one if the mother gave birth to twins at first birth. $\dot{z}$ indicates same-sex twins at first birth and $\bar{z}$ is our new twin instrument. Control variables are mothers' education (7 dummies), mothers' year of birth (43 dummies) and year fixed effects.

twins, and 0 for non-twin mothers. To construct $\bar{z}(\theta_{min})$, we derive θ_{min} as the θ which minimizes the overall F-statistic in a regression of $\bar{z}(\theta)$ on the covariates. A grid search delivers $\theta_{min} = -0.07$ for the whole sample and $\theta_{min} = 0.01$ for the sample of working mothers.

We study the effects of having more than two children on the probability of working and on annual labor income using our various instruments. The covariates are the same as in AE, but our sample size is 394,840 instead of 394,835. Nevertheless, we can virtually perfectly replicate their Table 7 for the worked-for-pay outcome in the first panel of *Table 4*. Column 1 reports a highly significant negative effect of -0.176 on the probability of working when we use OLS to estimate the fertility parameter. Using twins as instrument yields a coefficient of half that size, -0.081 (column 2). The IV effects differ little from each other when using $\ddot{z}$, $\dot{z}$, $\bar{z}(0)$, or $\bar{z}(\theta_{min})$ (columns 2–5). Note that the first stage F-statistic of 632 and 855 of the new instruments are much lower than those of the usual twin instruments $\ddot{z}$ and $\dot{z}$, but are still clearly above the rule of thumb value of 10 (Staiger

Table 4: Effect of having more than two children - US Census data

	OLS	2SLS				
		$\ddot{z}$	$\dot{z}$	$\bar{z}(0)$	$\bar{z}_t(\theta_{min})$	Same-sex
Worked for pay in last year (N=394,840)						
More than two children	-0.176***	-0.081***	-0.082***	-0.084**	-0.084***	-0.117***
	(0.002)	(0.014)	(0.017)	(0.034)	(0.031)	(0.025)
First Stage F-statistic		60,239	44,576	632	855	1,675
Selection on observables F-statistic		9.45***	3.85***	1.13	1.11	2.04*
Log(labor income) (N=220,502)						
More than two children	-0.353***	-0.072	-0.112**	-0.215*	-0.217*	-0.135
	(0.006)	(0.045)	(0.054)	(0.117)	(0.118)	(0.092)
First Stage F-statistic		35,754	25,484	292	280	841
Selection on observables F-statistic		8.65***	3.53***	0.72	0.72	1.75*

Notes: OLS and 2SLS estimates using data from the 1980 US Census. All models also include age, age at first birth, sex of the 1st child, sex of the 2nd child, and dummies for being black, hispanic, or of other race. Selection on observables F-statistic refers to the F-statistic of a regression of the respective instrument on the above covariates, except sex of the 1st and 2nd child for the same-sex instrument. θ_{min} equals -0.07 for worked in last year sample and 0.01 in income sample. Robust standard errors in parentheses. ***,**,* indicate significance at the 1%, 5%, and 10% level, respectively.

& Stock, 1997). Using the same-sex instrument (column 6), the coefficient is -0.117.

The last row in the first panel of *Table 4* reports the F-statistics of regressions of each respective instrument on the covariates to assess the importance of selection on observables (cf. *Table 2*). The overall F-statistic decreases from 9.45 to 3.85 when using the improved same-sex twin instrument, as compared to the overall twin instrument.[17] As in the Swedish data, our new instruments are the least correlated with the mothers' observable characteristics. Although there still seem to be small correlations with mother's age and age at first birth, the overall F-statistics of 1.13 and 1.11 are insignificant.

The second panel of *Table 4* shows the results for log-labor income as the dependent variable. In contrast to the employment indicator, we find large differences between the estimated effects using the usual instruments and those from using our new twin instruments. The absolute size of the coefficients increases with the share of monozygotic twin mothers in the instrument. The effect is lowest when using all twins (-7.19%), but almost triples (-21.47%) when using $\bar{z}(0)$, and reaches -21.70% when using $\bar{z}(\theta_{min})$. The

[17] To investigate whether this decline is systematic, we randomly exclude the fraction of doubtful (i.e., dizygotic) twins as in Section 2 and find the average F-statistic to be 6.87.

increase in the coefficients indicates that dizygotic twin mothers are a positively selected sample, which lead to an underestimation of the true effect. This was to be expected from the known relation between maternal characteristics (particularly, maternal age) and dizygotic twinning. For instance, women who earn more and/or have higher career preferences may also be more likely to postpone motherhood, which would increase the likelihood of dizygotic twinning. The first stage F-statistic behaves in the same way as for employment, being lowest for the new twin instruments. [18]

The estimate using the same-sex instrument of -13.53% is in between the estimates from the twin instruments, but statistically insignificant. The different effect size can be attributed to the identification of different local average treatment effects (Angrist et al, 1996).

5 Conclusions

Twin births are a popular instrumental variable for the endogenous fertility decision and family size. However, identification of causal effects might fail as having dizygotic twins is strongly related to mothers' age, height, weight, race, and the use of fertility treatments, such as in-vitro fertilizations. To overcome this, we provide a new instrument that corrects for the selection bias introduced by dizygotic twins, even if zygosity is unknown. The new approach depends on a parameter θ, which reflects the researcher's assumption about the strength of the relation between the structural error term and monozygotic twinning, relative to dizygotic twinning. In line with the medical literature, we find evidence for the exogeneity of monozygotic twinning (corresponding to $\theta = 0$). We could, however, also assume that monozygotic twins are not fully exogenous but are less endogenous than dizygotic twins ($\theta \in (-1, 1)$). In this case we propose to set the parameter θ to the value that minimizes the overall F-statistic from a regression of the new instrument on the observed variables under the assumption that the selection on observables is informative about the selection on unobservables. In contrast to the usual instruments (any twins

[18] Finally, note that AE use labor income in levels and include mothers with zero earnings instead of using log-labor income. For these reasons we have also used labor income in levels as outcome, resulting in very similar patterns as those for log-labor income.

and same-sex twins), we show using Swedish register data that the new instrument is completely unrelated to important pre-pregnancy outcomes.

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

References

Agüero, J & M Marks (2008) "Motherhood and female labor force participation: Evidence from infertility shocks", *American Economic Review*, vol 98, no 2, p 500–504.

Altonji, J G, T E Elder & C R Taber (2005) "Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools.", *Journal of Political Economy*, vol 113, no 1, p 151–184.

Angrist, J D & W N Evans (1998) "Children and their parents' labor supply: Evidence from exogenous variation in family size.", *American Economic Review*, vol 88, no 3, p 450–477.

Angrist, J D, G W Imbens & D B Rubin (1996) "Identification of causal effects using instrumental variables", *Journal of the American statistical Association*, vol 91, no 434, p 444–455.

Angrist, J D, V Lavy & A Schlosser (2010) "Multiple experiments for the causal link between the quantity and quality of children", *Journal of Labor Economics*, vol 28, no 4, p 773–824.

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

Becker, G S & H G Lewis (1973) "On the interaction between the quantity and quality of children", *Journal of Political Economy*, vol 81, no 2, p 279–88.

Bergh, C (2005) "Single embryo transfer: A mini-review", *Human Reproduction*, vol 20, no 2, p 323–327.

Black, S E, P J Devereux & K G Salvanes (2005) "The more the merrier? the effect of family size and birth order on children's education.", *Quarterly Journal of Economics*, vol 120, no 2, p 669–700.

Black, S E, P J Devereux & K G Salvanes (2007) "From the cradle to the labor market? the effect of birth weight on adult outcomes", *Quarterly Journal of Economics*, vol 122, no 1, p 409–439.

Black, S E, P J Devereux & K G Salvanes (2010) "Small family, smart family? family size and the iq scores of young men", *Journal of Human Resources*, vol 45, no 1, p 33–58.

Bomsel-Helmreich, O & W Al Mufti (2005) "The phenomenon of monozygosity: Spontaneous zygotic splitting", in Blickstein, I & L G Keith, eds, "Multiple Pregnancy: Epidemiology, Gestation, and Perinatal Outcome", CRC Press, p 94–101.

Bortolus, R, F Parazzini, L Chatenoud, G Benzi, M M Bianchi & A Marini (1999) "The epidemiology of multiple births", *Human Reproduction Update*, vol 5, no 2, p 179–187.

Braakmann, N & J Wildman (2014) “Fertility treatments and the use of twin births as an instrument for fertility”, *mimeo*.

Bronars, S G & J Grogger (1994) “The economic consequences of unwed motherhood: Using twin births as a natural experiment”, *American Economic Review*, vol 84, no 5, p 1141–1156.

Bulmer, M G (1976) “Is weinberg’s method valid?”, *Acta geneticae medicae et gemellologiae*, vol 25, p 25–28.

Butcher, K F & A Case (1994) “The effect of sibling sex composition on women’s education and earnings”, *Quarterly Journal of Economics*, vol 109, no 3, p 531–563.

Cáceres-Delpiano, J (2006) “The impacts of family size on investment in child quality”, *Journal of Human Resources*, vol 41, no 4, p 738–754.

Cáceres-Delpiano, J & M Simonsen (2012) “The toll of fertility on mothers’ wellbeing”, *Journal of Health Economics*, vol 31, no 5, p 752–766.

Conley, T G, C B Hansen & P E Rossi (2012) “Plausibly exogenous”, *Review of Economics and Statistics*, vol 94, no 1, p 260–272.

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

Fauser, B C, P Devroey & N S Macklon (2005) “Multiple birth resulting from ovarian stimulation for subfertility treatment”, *The Lancet*, vol 365, no 9473, p 1807–1816.

Fellman, J & A W Eriksson (2006) “Weinberg’s differential rule reconsidered”, *Human Biology*, vol 78, no 3, p 253–275.

Figlio, D, J Guryan, K Karbownik & J Roth (2014) “The effects of poor neonatal health on children’s cognitive development”, *American Economic Review*, vol 104, no 12, p 3921–55.

Hall, J G (2003) “Twinning”, *The Lancet*, vol 362, no 9385, p 735–743.

Hardin, J, S Selvin, S L Carmichael & G M Shaw (2009) “The estimated probability of dizygotic twins: A comparison of two methods”, *Twin Research and Human Genetics*, vol 12, no 1, p 79–85.

Jacobsen, J P, J W Pearce III & J L Rosenbloom (1999) “The effects of childbearing on married women’s labor supply and earnings”, *Journal of Human Resources*, vol 34, no 3, p 449 – 474.

Kulkarni, A D, D J Jamieson, H W Jones Jr, D M Kissin, M F Gallo, M Macaluso & E Y Adashi (2013) “Fertility treatments and multiple births in the united states”, *New England Journal of Medicine*, vol 369, no 23, p 2218–2225.

Lee, J (2008) “Sibling size and investment in children’s education: An asian instrument”, *Journal of Population Economics*, vol 21, no 4, p 855–875.

Lundborg, P, E Plug & A W Rasmussen (2014) “Fertility effects on female labor supply: Iv evidence from ivf treatments”, .

MacGillivray, I, M Samphier & J Little (1988) “Factors affecting twinning”, in MacGillivray, I, D M Campbell & B Thompson, eds, “Twinning and twins”, John Wiley & Sons.

Martin, J A, B E Hamilton & M J Osterman (2012) “Three decades of twin births in the united states, 1980-2009.”, *NCHS data brief*, , no 80, p 1–8.

Martin, J A & M M Park (1999) “Trends in twin and triplet births: 1980–97”, *National vital statistics reports*, vol 47, no 24.

Nevo, A & A M Rosen (2012) “Identification with imperfect instruments”, *Review of Economics and Statistics*, vol 94, no 3, p 659–671.

Pison, G & A V D’Addato (2006) “Frequency of twin births in developed countries”, *Twin Research and Human Genetics*, vol 9, no 2, p 250–259.

Reddy, U M, A M Branum & M A Klebanoff (2005) “Relationship of maternal body mass index and height to twinning”, *American College of Obstetricians and Gynecologists*, vol 105, no 3, p 593–597.

Rosenzweig, M R & K I Wolpin (1980a) “Life-cycle labor supply and fertility: Causal inferences from household models”, *Journal of Political Economy*, vol 88, no 2, p 328–348.

Rosenzweig, M R & K I Wolpin (1980b) “Testing the quantity-quality fertility model: The use of twins as a natural experiment”, *Econometrica*, vol 48, no 1, p 227–240.

Rosenzweig, M R & K I Wolpin (2000) “Natural ”natural experiments” in economics”, *Journal of Economic Literature*, vol 38, no 4, p 827–874.

Staiger, D & J H Stock (1997) “Instrumental variables regression with weak instruments”, *Econometrica*, vol 65, no 3, p 557–586.

Thurin, A, J Hausken, T Hillensjö, B Jablonowska, A Pinborg, A Strandell & C Bergh (2004) “Elective single-embryo transfer versus double-embryo transfer in in vitro fertilization”, *New England Journal of Medicine*, vol 351, no 23, p 2392–2402.

Tong, S & R V Short (1998) “Dizygotic twinning as a measure of human fertility”, *Human Reproduction*, vol 13, no 1, p 95–98.

Vere, J P (2011) “Social security and elderly labor supply: Evidence from the health and retirement study”, *Labour Economics*, vol 18, no 5, p 676–686.

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

Weinberg, W (1901) “Beiträge zur physiologie und pathologie der mehrlingsgeburten beim menschen”, *Pflügers Archiv - European Journal of Physiology*, vol 88, no 6-8, p 346–430.

Appendix

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

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

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

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

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

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

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

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

Publication series published by IFAU – latest issues

Rapporter/Reports

2016:1 Engdahl Mattias and Anders Forslund ”En förlorad generation? Om ungas etablering på arbetsmarknaden”

2016:2 Bastani Spencer, Ylva Moberg and Håkan Selin ”Hur känslig är gifta kvinnors sysselsättning för förändringar i skatte- och bidragssystemet?”

2016:3 Lundin Martin, Oskar Nordström Skans and Pär Zetterberg ”Kåren och karriären: Student-politiken som språngbräda”

2016:4 Brommesson Douglas, Gissur Erlingsson, Johan Karlsson Schaffer, Jörgen Ödalen and Mattias Fogelgren ”Att möta den högre utbildningens utmaningar”

2016:5 Egebark Johan ”Effekter av skatter på ungas egenföretagande”

2016:6 Mannelqivst Ruth, Berndt Karlsson and Bengt Järvholm ”Arbete och arbetsmarknad i sjuk-försäkringen”

2016:7 Rosenqvist Olof ”Rösträtt och ungdomars kunskap om politik”

2016:8 Lindgren Karl-Oskar, Sven Oskarsson and Mikael Persson ”Leder bättre tillgång till utbildning till ökat politiskt deltagande?”

2016:9 Moberg Ylva ”Är lesbiska föräldrar mer jämställda?”

2016:10 Hinnerich Björn T. and Jonas Vlachos “Skillnader i resultat mellan gymnasieelever i fristående och kommunala skolor”

2016:11 Engdahl Mattias ”Invandringens arbetsmarknadseffekter: lärdomar från den internationella litteraturen och svenska resultat”

Working papers

2016:1 Bastani Spencer, Ylva Moberg and Håkan Selin “Estimating participation responses using transfer program reform”

2016:2 Lundin Martin, Oskar Nordström Skans and Pär Zetterberg ”Leadership experiences, labor market entry and early career trajectories”

2016:3 van den Berg Gerard J., Lena Janys, Enno Mammen and Jens P. Nielsen ”A general semi-parametric approach to inference with marker-dependent hazard rate models”

2016:4 Egebark Johan “Effects of taxes on youth self-employment and income”

2016:5 Holmlund Helena “Education and equality of opportunity: what have we learned from educational reforms?”

2016:6 Rosenqvist Olof “Rising to the occasion? Youth political knowledge and the voting age”

2016:7 Lindgren Karl-Oskar, Sven Oskarsson and Mikael Persson “How does access to education influence political candidacy? Lessons from school openings in Sweden”

2016:8 Moberg Ylva “Does the gender composition in couples matter for the division of labor after childbirth?”

2016:9 Hinnerich Björn T. and Jonas Vlachos “The impact of upper-secondary voucher school attendance on student achievement. Swedish evidence using external and internal evaluations”

2016:10 Farbmacher Helmut, Raphael Guber and Johan Vikström “Increasing the credibility of the twin birth instrument”

Dissertation series

2015:1 Egebark Johan “Taxes, nudges and conformity. Essays in labor and behavioral economics”

2015:2 Blind Ina “Essays on urban economics”