Göteborg Papers in Economic History

No. 22. April 2017 ISSN: 1653-1000

An introduction to using twin births as instrumental variables for sibship size

Stefan Öberg

DEPARTMENT OF ECONOMY AND SOCIETY

UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

UNIT FOR ECONOMIC HISTORY

An introduction to using twin births as instrumental variables for sibship size*

Stefan Öberg

stefan.oberg@econhist.gu.se

Abstract: Some families who experience a twin birth get one more child than they had intended and planned for. This is the reason why twin births are used to create instrumental variables (IVs) for the number of children in a family. In this chapter I introduce IV techniques in general and the use of twin births for IVs in particular. IVs based on parity-specific twin births can indeed, under certain circumstances, be valid and reliable. In this chapter I discuss what these circumstances are. I rely heavily on the work by Joshua Angrist and coauthors. In contrast to them I argue that it is important to recognize that IVs based on parity-specific twin births have "heterogenous treatment effects", meaning that it is only for some families that the twin birth leads to an unintended and unplanned birth. Recognizing this highlights a few assumptions that are not always thoroughly acknowledged in previous research. We, for example, need to make assumptions about the possibility of unintended single births and the families experiencing these. It is also the case that including families that have not yet reached (or surpassed) their desired number of children when using IVs based on parity-specific twin births will lead to estimates that are biased towards zero. Most importantly we need to reduce the claims of estimating generalizable, causal effects when using twin birth instrumental variables.

JEL: C26, C36, J13

Keywords: instrumental variables, twins, twin births, family size, exogenous variation, causal estimation, sibship size

ISSN: 1653-1000 *online version*
ISSN: 1653-1019 *print version*

University of Gothenburg
School of Business, Economics and Law
Department of Economy and Society
Unit for Economic History
P.O. Box 625
SE-405 30 GÖTEBORG
http://es.handels.gu.se/english/units/unit-for-economic-history/

* The author gratefully acknowledges financial support from the Jan Wallanders och Tom Hedelius foundation in the form of a Wallander PostDoc (W2014-0396:1). The author also thanks Damian Clarke and Sonia Bhalotra for comments on an earlier draft of the paper.

1. Introduction

Studies investigating how children are affected by their sibship size, i.e. the number of children in the family, face a serious challenge because the sibship size is endogenous in the model. One important reason for this endogeneity is that there are unobserved differences between parents that chose to have different number of children that could also be related to the life chances of their children. The most common method to solve the problems of endogeneity in this type of studies has been to use instrumental variables. Using parity-specific twin births to create these instrumental variables is considered the "gold standard" for this method. Because twin births are (assumed to be) random they create a "natural experiment" situation. These instrumental variables have therefore this far been considered relatively unproblematic. Bhalotra and Clarke (2016) have recently shown convincingly that we have to question the validity of instrumental variables based on twin births for many applications. The challenge they raise is serious enough to the method that we need to reevaluate the "gold standard" method for this line of research.

To make this possible for a, hopefully, wider than usual audience I try to provide a non-technical introduction to instrumental variables regressions in general and instrumental variables based on twin births in particular. It is a basic introduction from one non-expert to other non-experts, so that it risk coming across as self-evident or overly simplistic to some readers.

My introduction relies heavily on the work of Joshua Angrist and coauthors (Angrist, Imbens, and Rubin 1996; Angrist and Krueger 2001; Angrist and Pischke 2009). While I otherwise follow them closely I do make an adjustment when I apply their framework to analyze the use of twin births for instrumental variables. Angrist and Pischke (2009) argue that instrumental variables based on twin births are an exception among instrumental variables. Their argument for why they are a special case is that the "assignment to the treatment"—i.e. experiencing a (parity-specific) twin birth—corresponds completely with the "treatment"—having an "extra" child (p. 160–161). If this is correct it reduces the number of assumptions needed for using twin births for instrumental variables, and it also improves our opportunities to interpret the estimated effect as generalizable, causal effects.

I will argue that we should not view instrumental variables based on twin births as a special case, but rather acknowledge that they have "heterogenous treatment effects". In practice this means that I argue that we should not consider the "extra" child born as the treatment. We are not interested in using twin births for instrumental variables because they increase the number of children born, but rather because they sometimes lead to a randomly assigned and *unintended* increase of the number of children born. I argue that we should therefore consider such unintended births as the treatment rather than all the treatments. This argument is in line with Angrist and coauthor's definitions and explanations of instrumental variables in general and only deviates in regard to the specific interpretation of how we can use twin births for instrumental variables. I argue that if we consider unintended births as being the treatment in the twin birth instrumental variable case, this in some ways facilitates the interpretation and evaluation of the instruments. But, it also highlights a couple of previously overlooked assumptions that are necessary for the analyses. We, most importantly, must consider if parents that experience an unintended single birth could be different from parents that experience an unintended birth because of a twin birth. If my interpretation of instrumental variables based on twin births is the more reasonable, we also have to reduce the claims regarding the estimated effect. All is of course not lost, but the estimated effects should be considered to be less generalizable.

2. Presentation of the simulated data used for illustrations

To illustrate a few key points about using twin births for instrumental variables I have run some simple simulations and present the results below. I simulated a very large population with at least one birth (N = 1’000’000).[1] These families were given a fixed desired number of children randomly decided within a distribution. A fixed desired number of children is likely a too restrictive assumption, even if it is plausible that most families have some kind of idea of the number of children that they would like and how many would be less convenient. Anyway, when we use twin births for instrumental variables for sibship size we assume that families have a sufficiently fixed desired number of children so that some (especially twin) births can result in an “unwanted” birth.[2]

We rarely, especially for historical populations, know anything about parents’ desired number of children. For the simulation I have therefore had to assume hypothetical distributions. I vary these distributions to illustrate how the results depend on the

[1] The script for running the simulation using R (R Core Team 2016) is available from the author.

[2] I will call these births “unwanted” in this text. This pointed term is merely intended to make my argument as clear as possible. The children in reality do not have to actually be unwanted in their families for the instrumental variables based on twin births to work. What is needed is that some parents have one more child than they had intended and planned for.

distribution, i.e. on the investigated population and time period. I use the distributions of the observed number of children from four studies that investigate the association between sibship size and child outcomes and that, more importantly, present the distribution of observed number of children in the paper (Black, Devereux, and Salvanes 2005, tab. II; Åslund and Grönqvist 2010, tab. 1; Stradford, van Poppel, and Lumey 2017, tab. 1; Roberts and Warren 2017, tab. 3). This is far from perfect but does provide some empirical basis for the shapes of the distributions.

All families in the simulated populations, as mentioned, have at least one birth. They then go on to have another birth until they reach their desired number of children. Each birth has a small and constant chance of being a twin birth ($p=0.0175$). (For simplicity, I do not include any other types of multiple births.) Because of twin births some families will exceed their desired number of children. These simulated data are of course a highly simplified version of reality. They are solely intended to illustrate what twin birth instrumental variables capture in situations when they work well. The twin births would in this simulated case by construction make valid instrumental variables.

Most of the values presented from the simulation are determined by the shape of these distributions together with the chance for twin births. I still repeated the simulation to investigate how much the values would vary by chance. They turned out not to vary much at all. I therefore only present the median value from the simulations in the tables below.

3. The challenge of endogeneity when studying the effect from sibship size on children

Investigations into how children are affected by the sibship size, i.e. the number of children in the family, have a long history in both social sciences and public health research. A large number of studies have found negative associations between sibship size and different outcomes of the children, for example with regard to their education and social mobility (e.g. Blau and Duncan 1978, chap. 9; Blake 1981; 1985). These negative associations have later been confirmed for many populations in high-income countries (Park 2008; Xu 2008; Kalmijn and Werfhorst 2016). Negative associations have also been shown for a wide range of other outcomes, including the height of the children (see e.g. Öberg 2015 and the references therein).

Understanding the causes underlying the negative associations will help us understand a range of other things. Influences from the sibship size could, for example, be part of the mechanism for the transmission of social (dis-)advantage. How children are affected by their sibship size also provide insights into the costs of raising children and how families manage these costs in different contexts and times. This has important implications for, for example, theories on the fertility decline, investments in human capital and inter-generational transfers of resources.

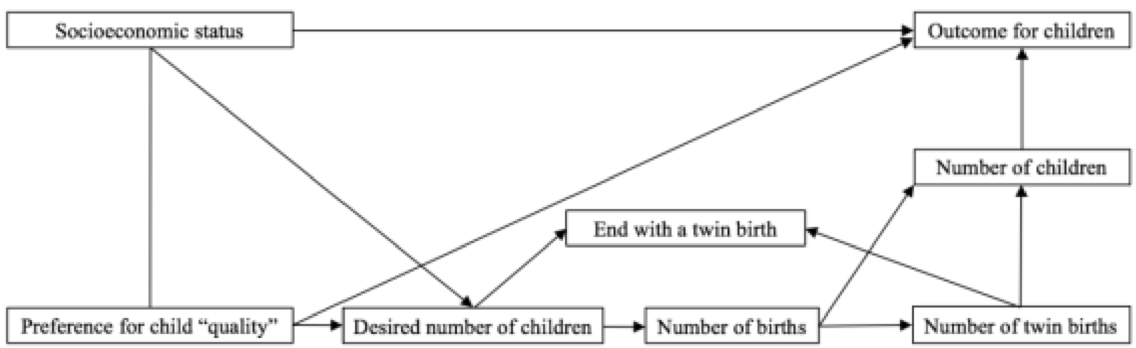
The most commonly used framework for understanding the negative associations is the resource dilution hypothesis (Blake 1981; 1985; Gibbs, Workman and Downey 2016). This provides an intuitive explanation by claiming that the cause behind the negative associations between sibship size and child outcomes is that parental resources are limited and hence diluted in families with many children. The parental resources include material resources but also time, energy, patience etc. These resources are limited and parents will not be able to counteract scarcity across all aspects when they have many children to care for.

Resource dilution is also one of the underlying assumptions of the theoretical model developed by Gary Becker and colleagues on how families invest in their children (e.g. Becker 1993). The model posits that parents make a choice on how many children to have and how much to invest in them. If they have many children they can invest less in each, if they have fewer children they can invest more. The more the parents invest in their children, the more they improve their "quality". The parents thus have to make a trade-off between the *quantity* and *quality* of their children.

The complementary explanation for the negative associations between sibship size and child outcomes, besides resource dilution, is that parents who chose to have different number of children are different also in other ways. If there are differences between parents with different number of children this can create problems for our analyses. If these differences influence both the number of children *and* the outcomes for the children, they will act as confounders of the estimated association between sibship size and the outcome. The parents with different number of children can be different in observable ways. An important example of this is the socioeconomic gradients in fertility that are observed in many populations (Dribe et al. 2017). Such gradients could mean, for example, that children with many siblings also have parents with less education and lower incomes. To investigate the association between the sibship size and the outcomes for the children, in this example, we would therefore have to adjust our estimates for indicators the parents' education and income.

The parents with different number of children can also be different in unobservable ways, such as in their ability or preferences. Parents who are better able to care for children and have a preference for spending more on their children could, for example, be more likely to have a larger number of children and also better outcomes for their children. Or, it could be that parents with a preference for child "quality" desire fewer children to be able to invest more in each. This is what is assumed in the theoretical quantity-quality framework of Becker (1993). These unobservable differences will create problems for our analyses if they confound the association between sibship size and the outcomes for the children. It is a very reasonable assumption that they will work as confounders (Figure 1), and so we need to take this into consideration when investigating this issue.

FIGURE 1. *Graphical representation of the association between the number of children in the family and the outcome for the children*



There are several reasons for why we can't think of the number of children as being determined independently of the outcomes for the children. I will focus on differences between parents confounding the association because this is likely to be the case empirically and because it is assumed in one of the theoretical frameworks tested.

Regardless of where we will end up with our analysis in regard to method, we always need to start by considering what plausible factors could influence the association in the case we are studying and how. There will always be a very large number of possible factors potentially influencing the outcome. What we are looking for are factors that could influence both the explanatory variable of interest *and* the outcome studied. Examples of such factors when studying the effect from sibship size on child outcomes are the socio-economic status of the family and the parents' preferences for child "quality" (Figure 1).

The next step is to consider if we can observe and include all relevant factors in the model. If this is possible we proceed and estimate the effect using ordinary least squares regression. If it is not possible to observe and/or include all relevant factors in the model we need to find a way to take this unobserved factor(s), i.e. omitted variable(s), into account. As long as there are differences between parents with different number of children that we do not (perfectly) adjust for, we cannot estimate the causal effect from the sibship size on the outcome from just comparing the outcomes for children from families with different number of children.

If, for example, parents with a strong preference for child "quality" have fewer children and also invest more in each child this will create a spurious negative association between the number of children and child outcomes. The results are biased in the same way as from omitting any other relevant explanatory variable. Because we have an omitted variable that affects both the number of children and the studied outcome, the variable measuring the number of children will be correlated with the residuals from the regression. Because the variable is correlated with the residuals it cannot be treated as an exogenous factor. Problematic variables, such as the number of children, are therefore called "endogenous" variables. If we have endogenous variables the results from, for example, ordinary least squares regressions will be biased. We therefore have to find a way to deal with the endogenous variable(s) to be able to estimate the actual, causal effect from sibship size on child outcomes. The most commonly used such method is instrumental variables regressions.

4. An introduction to instrumental variable techniques

The most commonly used solution for estimating unbiased, including unconfounded, effects when the explanatory variable is endogenous is instrumental variable techniques (for an introduction, see: Angrist and Pischke 2009, chap. 4; see also, for example: Angrist, Imbens, and Rubin 1996; Angrist and Krueger 2001; Murray 2006; Bollen 2012). These techniques attempt to isolate "exogenous" variation in the endogenous variable and analyze its association with the outcome using only this variation. In practice this means using only the variation of the endogenous variable that is related to something that, in turn, is not related to the outcome studied. This "something" is the instrument or instrumental variable. It needs to be, in itself, not related to the outcome studied nor related to other aspects that cause a spurious association. This is what makes the instrumental variable "valid". The instrument needs to be, in contrast, fairly closely related to the endogenous variable. This is what makes the instrumental variable "relevant" or "informative".

The most difficult requirement on an instrumental variable is that it should not be associated with the outcome we are studying; stated differently it should be "exogenous" to the outcome. The instrumental variable can, self-evidently, not have any direct effect on the outcome studied. It is therefore possible to exclude (and should be excluded) from the model of the outcome that we estimate. This requirement on the instrumental variable and assumption in the method is therefore called the "exclusion restriction". We can, sometimes, find an instrumental variable that doesn't have any direct influence on the outcome. But, what is further required is that it should not be related to anything unobserved that is also related to the outcome. There should, in other words, not be any unobserved factors confounding the association between the instrument and the outcome. There can be observable factors confounding the association if we can (perfectly) adjust our models for these factors. If we (perfectly) adjust the models for these observables, the instrumental variable will be conditionally independent of the outcome. One way of achieving this is if the assignment to the event underlying the instrumental variable is random. In such cases we trust that the independence assumption holds *on average and in large samples*.

We also need the instrumental variable to have an influence on the potentially problematic, "endogenous", variable. If the instrumental variable does not influence the endogenous variable, it is not "relevant". This might seem like a trivial requirement but given the difficulty in finding instrumental variables we need to check that what we use actually have a substantial effect on the endogenous variable. We can check this, for example, by regressing the endogenous variable on the instrumental variable using ordinary least squares. There is then a rule-of-thumb that the *t*-statistic of the coefficient on the instrumental variable should be equal to ten or higher for the instrument to have a sufficiently strong and systematic impact on the endogenous variable to work well (Staiger and Stock 1997). Instrumental variables that are only weakly associated with the endogenous variable are called "weak" instruments and can create a whole range of problems for the analysis (e.g. Stock, Wright and Yogo 2002; Murray 2006).

The requirements that the instrumental variable should have a (fairly) strong effect on the endogenous variable while being (conditionally) independent of the outcome are very difficult in practice. What social scientists, including economists, oftentimes end up using are random events or decisions that are not made by the people studied but still affect them. The most common and best known example of an instrumental variable in the literature studying the influences from sibship size on child outcomes is that of twin births. This has in itself a long history (starting with Rosenzweig and Wolpin 1980) and is indeed, under certain circumstances, a valid and reliable instrument. Below I will discuss what these circumstances are.

5. An introduction to using twin births for instrumental variables

The idea behind using twin births for instrumental variables is that some of the families which have a (parity-specific) twin birth are getting an "extra" child that they would not otherwise have had. Because twin births are (assumed to be) random events, we get a situation where a group of randomly chosen families are assigned an "extra" child. This is how we can argue that the twin births create an experiment-like situation.[3] Because nature, through human biology, creates this situation it is included among so-called "natural experiments" (Rosenzweig and Wolpin 2000).

For instrumental variables based on twin births the exclusion restriction means that experiencing a (parity-specific) twin birth shouldn't in itself influence the other children in the family. This is, in many cases, plausible. There should also be nothing that influences both the chance of a twin birth and the studied outcome for the children. This is possible but a more difficult requirement than the first. If there are such unobserved factors this will lead to the instrument being only conditionally exogenous. Conditionally independent instrumental variables lead to biased results if we cannot (perfectly) adjust our model for the factors that do influence both the chance of a twin birth and the studied outcome for the children. Instrumental variables based on twin births are assumed to be independent of the outcome because they are assumed to be random events. Because the twin births are random events we can assume that there will, *on average and in large samples*, be no systematic differences between those indicated and not by the instrumental variable, i.e. experiencing a (parity-specific) twin birth or not. We can for example and importantly assume that there should be no association between experiencing a (parity-specific) twin birth and the desired number of children. This is why we can estimate the un-confounded, causal effect of sibship size on the outcome by comparing the average outcome and number of siblings for children in families who experience a (parity-specific) twin birth with children in families that do not.

To uphold the assumption of random assignment of twins it is important that we construct the instrumental variable in a way so that the twin birth can indeed be seen as a

[3] In reality the design is, of course, not exactly as a randomized, controlled experiment, and using instrumental variables adds another level of assumptions to the implied behavioral model (Rosenzweig and Wolpin 2000).

random event. What is important to keep in mind is that it is for each birth that the twin/not twin event can be treated as a random event. We can therefore, for example, not use an indicator for families that experienced any twin birth as an instrumental variable. The chance of this happening is, quite naturally, increasing with each birth and so this indicator will be positively associated with the desired number of children.

For the instrumental variables created from twin births not to be associated with the desired number of children we need to create them based on parity-specific twin births. We then use them in "*n*+ samples". This means that we use a twin birth at a specific parity as the instrument and include the families with that many *or more* births in the analysis. Because the twins have some special features (e.g. Silventoinen et al. 2013) they are themselves excluded from the analysis. We also don't want to include other later born children since they are only present in the larger families. When we use twin births to create instrumental variables we therefore in practice study how the older, earlier born, children are affected by an exogenous increase in the number of younger siblings. We do this in families with more than two births. If we, for example, use a twin birth as the second birth to create our instrumental variable we study only the first-born children in families with at least two births. We also need to make sure that we only include families in our sample that have reached (or surpassed) their desired number of children. Otherwise the estimated effect from sibship size on the child outcomes will be biased towards zero.

5. Understanding twin births as instrumental variables through a counterfactual analysis

To further investigate if the necessary assumptions for an instrumental variable are met it is useful to use the counterfactual framework applied in, for example, Angrist, Imbens and Rubin (1996). This helps highlighting the assumptions required if we want to use instrumental variables based on twin births. The first thing that I think this framework helps to highlight is that there is a difference between the random event that creates the instrumental variables, the twin births, and what we intend for these instrumental variables to capture, exogenous increases of the number of children in the family.[4] What a twin birth does is to make an unexpected increase in the number of children. Some families will have always wanted to have another two or even more children and just have them quicker than expected. For these families the twin birth thus does not lead to an exogenous increase of the number of children.

[4] My analysis and conclusion here differs from that of Angrist and Pischke (2009).

A twin birth can only result in an "unwanted" birth if we assume that each family has a set desired number of children. This is most likely a too restrictive assumption, but it is plausible that many families have some kind of idea of the number of children that is convenient and what is less convenient.

By separating the random assignment to the instrumental variable—a parity-specific twin birth—from its consequence—an exogenous increase of the number of children—it becomes easier to evaluate the necessary assumptions. Families either experienced a (parity-specific) twin birth or not, and this resulted in an extra child being born that was either "wanted" (endogenous) or "unwanted" (exogenously increasing the number of children). To evaluate the use of (parity-specific) twin births as instrumental variables we need to consider differences between the four groups with the four different possible combinations of these outcomes. These four groups will each include two groups of families with different characteristics. The analysis depends on that some of these groups do not exist or that there are no systematic unobserved differences between these groups.

Firstly, we have the "compliers" (see Angrist, Imbens, and Rubin 1996 for further discussion of these terms). The compliers are the families who had no twin birth (at the studied parity) and who get exactly as many children as they wanted, and the families who had a twin birth (at the studied parity) and therefore had one more child than they wanted. This is the group for which the instrumental variable works as intended.

Secondly, we have, besides the compliers, three groups that deviate in different ways from the expected pattern. There are the "never-takers", which in the twin instrumental variable case correspond to the families that get the exact number of children they wanted regardless of whether they had a twin birth or not (at the studied parity).[5] Experiencing twin births or not should (conditionally) not be a useful predictor of the outcome for the children, i.e. there should be no systematic, unobserved differences between families that experience a twin birth and those that do not. This is the exogeneity assumption that makes the twin birth instrumental variables valid, and has, prior to Bhalotra and Clarke (2016), mostly been seen as a relatively unproblematic assumption.

For the other groups we need to allow for "unwanted" births also among single births. One of the groups including "unwanted" single births are the "always-takers", which in the twin birth instrumental variable case correspond to all the families that experience an

[5] They are "never-takers" of the "treatment" of having an "unwanted" birth.

"unwanted" birth, single or twin. For this group the exogeneity assumption requires that there should not be any differences between families that experience an "unwanted" birth because of a twin birth and families that experience an "unwanted" birth as a result of a single birth. This needs to be true for the twin birth instrumental variables to be valid. It is not difficult to think of reasons why this might not be true, but the assumption could be sufficiently accurate empirically. This assumption is, anyway, much less self-evident than those mentioned above. An alternative assumption for making the twin birth instrumental variables valid would be that there are no "unwanted" single births at all. This is an even more strenuous assumption than what was just discussed.

The last category we evaluate is the "defiers", which in the twin instrumental variable case correspond to families that "defy" their classification in the instrumental variable. The "defiers" include families that change their mind about wanting a child when they have a single birth rather than a twin birth. They also include families with parents that are fundamentally changed in their preferences and behaviors regarding how many children they want and how they treat their children by the experience of having a twin birth. This last group is less implausible than the first, but both can be expected to be small if they exist.

5. How a two-stage least squares instrumental variable regression works

The most intuitive, and commonly used, method for using instrumental variables in regressions is "two-stage least squares". For this we formulate two models: one modelling the influence from the problematic, endogenous explanatory variable on the outcome studied, and another modelling the variation in the problematic, endogenous explanatory variable in itself. Only the model of the problematic, endogenous explanatory variable includes the instrumental variable(s). The assumption is, as mentioned, that a (parity-specific) twin birth can contribute to predicting the number of children in a family, but does not influence the studied outcome for the older children in these families.

The first stage of a two-stage least squares regression is that we regress the endogenous variable on the instrumental variable.[6] If there are other explanatory variables in the model of the outcome, other than the problematic variable, we include also those in this first stage model.[7] In the second stage, we replace the observed number of children with the

[6] The problematic, endogenous explanatory variable is thus the dependent variable in this model.

[7] The results from this first stage regression is, as mentioned, where we can check that the instrument is relevant and not "weak". We do this by looking at the size of the coefficient on the instrumental variable and its *t*-statistic.

predicted values from the first stage regression. When we use (parity-specific) twin births to create instrumental variables these are, quite naturally, binary. Either the family had a twin birth at the studied parity, or not. I will explain how the method works while assuming that there are no other explanatory variables, because it makes it so much easier to understand the method. The principles explained are not changed if we do include other explanatory variables in the first stage.

For the second stage regression we, as mentioned, use only the variation in the number of children that is related to the binary instrument. In a simplified case, this, in practice, means that the variation in the number of children is reduced to a binary variable as well.[8] All families that experienced a twin birth at the studied parity get the same predicted value from the first stage. In the simplified case this corresponds to the constant (i.e. the conditional average number of children in the sample) plus the coefficient on the instrumental variable. The predicted value for families that did not experience a twin birth at the parity is, in the simplified case, just the constant. The difference in the average number of children between families experiencing a (parity-specific) twin birth and not, i.e. the first-stage coefficient on the twin birth variable, is the only variation in the number of children that is used to estimate its effect on the outcome. The instrumental variable regression can therefore, in this simplified case, be intuitively understood as estimating a regression line from the differences in means of two variables for two groups, i.e. drawing a line between two points in a scatter plot. We estimate the effect from the number of children on the outcome by how much larger the number of children, on average, are for families experiencing a (parity-specific) twin birth compared to those that do not, and by the average difference in the outcome for the children in these two groups.[9]

[8] In practice we will, almost always, have also other explanatory variables in the first stage model of the endogenous variable. These will also influence the predicted values through their estimated coefficients. When we include also other explanatory variables in the first stage regression the predicted values are a so-called “linear combination” of the included explanatory variables and the instrumental variable. This means that the predicted values are a sum of the observed variable values multiplied by their respective coefficients. The reason for why this does not lead to issues of multicollinearity when we include this prediction in the model of the outcome, which also includes the other explanatory variables, is the instrumental variable. This creates variation in the predicted values that are not part of the other explanatory variables. This is another reason for why we need the instrumental variables to be not “weak”, i.e. add a substantive amount to the variation in the predicted values.

[9] We use instrumental variable *regressions,* rather than estimating the effect based on the group means (which is called the Wald estimator), because there are almost always relevant observable factors we should adjust the estimates for besides the instruments. This extension does not change the intuition of the method as I try to outline it here.

The twin birth instrumental variables will, in practice, indicate families with different number of children. This heterogeneous group is compared to the equally heterogeneous families that did not experience a twin birth at the studied parity. The heterogeneity of the families indicated by the instrumental variable reduce the difference between them and the rest of the sample. We can still get an accurate estimate of the effect because both the difference in the problematic explanatory variable—sibship size—and in the outcome, are reduced. As long as the instrumental variable is valid, there should be nothing else creating systematical differences between the families indicated by the instrumental variable and not, other than that some of the indicated families experienced an "unwanted" birth because of the twin birth. We can therefore still estimate the slope of the regression line from the two points in the scatter plot because they close in on each other along the line that we want to estimate the slope of. The estimated slope should be accurate as long as the assumptions hold, but the estimation risks becoming volatile when the points are closer together. This is why instrumental variable regressions become more sensitive to violations of the assumptions when the instrumental variable is weak, i.e. doesn't have a substantive (and systematic) effect on the endogenous variable.

Thinking about the instrumental variable regression in this way makes it more similar to methods we all know and have used. It might therefore be easier to see why there can be no systematic differences between the groups that could confound the observed difference in the outcome. It also becomes clear why the method does not work well when the difference in, for example, the number of children is small. If this difference is small, we don't have much variation in the explanatory variable between the groups we are comparing and so the results become less robust, just as for other methods, such as ordinary least squares regressions.

6. Illustrating the heterogenous treatment effect of instrumental variables based on twin births through the first-stage regression coefficient

We assume that the families who experience a (parity-specific) twin birth are not in any systematic way different from the families that do not experience this. We, for example and importantly, assume that these two groups of families do not desire different number of children. We should therefore not expect the families experiencing a (parity-specific) twin birth to proceed to higher birth orders because of the twin birth. This would make these parents being systematically different because of the twin births. This would, in turn, make it relevant to include an indicator for these families to adjust our models of the outcome, thus violating the exclusion restriction (making the instrumental variable not valid).

Because we assume no systematic differences between families who do or do not experience a (parity-specific) twin birth, the difference in number of children between the groups should capture only the exogenous increase in *some* families, i.e. the "unwanted" births that exogenously increase the number of children.

The size of the first stage coefficient corresponds to the share out of the families that experienced a twin birth at the studied parity, that wanted *n* children but got *n+1* because of the twin birth, i.e. an "unwanted" birth (Table 1). The results in Table 1 shows that the twin birth instrumental variables work as intended when they are valid (as they are by construction in the simulation). The first stage coefficient captures the expected number of "unwanted" children born to a family experiencing a twin birth (at the studied parity). The share of the twin births that are "unwanted", quite naturally, depends strongly on the hypothetical distribution of the desired number of children. Both the average and the shape of this distribution influence the shares "unwanted" at the different birth orders. The coefficient will be smaller in higher fertility desire populations because a larger share of the parents wants *n+1*, or even more, children (Table 2).

Even if the size of the coefficient varies there is always a substantial difference in the number of children between families who do or do not experience a (parity-specific) twin birth. The coefficient from the first-stage regression is therefore often in the middle or in the upper half of its range between zero and plus one, and with a *t*-statistic well above ten (Table 1). And so, even if the twin births explain only a miniscule part of the variation in family size, these instruments are not "weak".

Because we intend for the twin birth instrumental variable to capture "unwanted" births we should expect the first-stage coefficient to always be between zero and plus one. A coefficient outside this range is a clear indication that the instrument is capturing some other variation than the intended. The first-stage coefficient is around 0.7–0.8 in present-day populations (Bhalotra and Clarke 2016) which would indicate that about that share of the twin births leads to an "unwanted" increase of the number of children. This is high but not in itself implausible. But, I would still argue that it is surprisingly high if we allow for the possibility of "unwanted" single births. "Unwanted" single births will reduce the difference in final number of children between the groups of parents that do and do not experience the (parity-specific) twin birth, and thus reduce the size of the first-stage coefficient.

TABLE 1 *Results regarding basic properties of instrumental variables based on parity-specific twin births from analyses of simulated populations*

	b, first stage	Share of the indicated twin births that are "unwanted"	Share of population wanting exactly *n* children	Share wanting *n* + 2 children or more	Share of population included in *n* + sample	R^2, first stage	*t*, first stage
Distribution based on Black, Devereux, and Salvanes (2005)							
Twin as 2nd birth	0.494	0.503	0.409	0.146	0.814	0.0042	58.4
Twin as 3rd birth	0.643	0.647	0.266	0.047	0.402	0.0086	59.2
Twin as 4th birth	0.673	0.676	0.099	0.016	0.141	0.0094	36.5
Twin as 5th birth	0.659	0.664	0.031	0.006	0.045	0.0083	19.4
Distribution based on Åslund and Grönqvist (2010)							
Twin as 2nd birth	0.507	0.516	0.453	0.145	0.876	0.0043	61.8
Twin as 3rd birth	0.660	0.666	0.287	0.048	0.422	0.0084	59.7
Twin as 4th birth	0.663	0.668	0.097	0.018	0.140	0.0074	32.4
Twin as 5th birth	0.622	0.627	0.030	0.007	0.046	0.0057	16.2
Distribution based on Stradford, van Poppel and Lumey (2017)							
Twin as 2nd birth	0.390	0.400	0.269	0.233	0.678	0.0009	25.2
Twin as 3rd birth	0.433	0.444	0.181	0.131	0.408	0.0013	22.7
Twin as 4th birth	0.429	0.442	0.101	0.076	0.227	0.0014	17.7
Twin as 5th birth	0.422	0.429	0.056	0.044	0.127	0.0016	14.3
Distribution based on Roberts and Warren (2017)							
Twin as 2nd birth	0.328	0.340	0.280	0.346	0.840	0.0006	22.3
Twin as 3rd birth	0.383	0.394	0.218	0.203	0.557	0.0009	22.6
Twin as 4th birth	0.407	0.420	0.143	0.114	0.339	0.0012	19.9
Twin as 5th birth	0.434	0.442	0.089	0.066	0.197	0.0015	17.3

The first-stage regression coefficient will only correspond to the share of the twin births that are “unwanted” if we study only families that have all reached (or surpassed) their desired number of children. The first-stage coefficient will therefore also depend on how long time has passed since the twin birth. If not that much time has passed the parents who always wanted more children will not have had time to realize this. The difference between parents who did and did not experience a parity-specific twin birth will therefore be larger if less time has passed since the twin birth. The argument I have outlined about how the first-stage coefficient (almost exactly) captures the share of the births that are "unwanted" relates to when all families have reached or surpassed their desired number of children. The conceptual way of thinking about the first-stage regression coefficient presented above still holds, but they will not correspond exactly to the share of the births that are "unwanted".

TABLE 1 *The accuracy of the twin birth instrumental variables*

A. Distribution based on Black, Devereux, and Salvanes (2005)

Twin as 2nd birth, *b*, first stage = 0.494, Accuracy = 98.3%		Twin birth at the studied parity: No	Yes	Twin as 3rd birth, *b*, first stage = 0.643, Accuracy = 98.8%		Twin birth at the studied parity: No	Yes
"Unwanted birth"	No	0.9739	0.0085	"Unwanted birth"	No	0.9764	0.0060
	Yes	0.0087	0.0088		Yes	0.0062	0.0113
Twin as 4th birth, *b*, first stage = 0.673, Accuracy = 98.9%		Twin birth at the studied parity: No	Yes	Twin as 5th birth, *b*, first stage = 0.659, Accuracy = 98.8%		Twin birth at the studied parity: No	Yes
"Unwanted birth"	No	0.9770	0.0056	"Unwanted birth"	No	0.9767	0.0058
	Yes	0.0057	0.0118		Yes	0.0059	0.0117

B. Distribution based on Åslund and Grönqvist (2010)

Twin as 2nd birth, b, first stage = 0.507, Accuracy = 98.3%		Twin birth at the studied parity: No	Yes	Twin as 3rd birth, b, first stage = 0.660, Accuracy = 98.8%		Twin birth at the studied parity: No	Yes
"Unwanted birth"	No	0.9742	0.0085	"Unwanted birth"	No	0.9768	0.0059
	Yes	0.0083	0.0090		Yes	0.0057	0.0117
Twin as 4th birth, b, first stage = 0.663, Accuracy = 98.8%		Twin birth at the studied parity: No	Yes	Twin as 5th birth, b, first stage = 0.622, Accuracy = 98.7%		Twin birth at the studied parity: No	Yes
"Unwanted birth"	No	0.9768	0.0058	"Unwanted birth"	No	0.9761	0.0065
	Yes	0.0057	0.0117		Yes	0.0064	0.0110

C. Distribution based on Stradford, van Poppel, Lumey (2017)

Twin as 2nd birth, b, first stage = 0.390, Accuracy = 97.9%		Twin birth at the studied parity: No	Yes	Twin as 3rd birth, b, first stage = 0.433, Accuracy = 98.1%		Twin birth at the studied parity: No	Yes
"Unwanted birth"	No	0.9722	0.0105	"Unwanted birth"	No	0.9729	0.0097
	Yes	0.0103	0.0070		Yes	0.0096	0.0078
Twin as 4th birth, b, first stage = 0.429, Accuracy = 98.1%		Twin birth at the studied parity: No	Yes	Twin as 5th birth, b, first stage = 0.422, Accuracy = 98.0%		Twin birth at the studied parity: No	Yes
"Unwanted birth"	No	0.9729	0.0098	"Unwanted birth"	No	0.9728	0.0100
	Yes	0.0096	0.0077		Yes	0.0098	0.0075

D. Distribution based on Roberts and Warren (2017)

Twin as 2nd birth, b, first stage = 0.328, Accuracy = 97.7%		Twin birth at the studied parity: No	Yes	Twin as 3rd birth, b, first stage = 0.383, Accuracy = 97.9%		Twin birth at the studied parity: No	Yes
"Unwanted birth"	No	0.9711	0.0116	"Unwanted birth"	No	0.9721	0.0106
	Yes	0.0114	0.0060		Yes	0.0104	0.0069
Twin as 4th birth, b, first stage = 0.407, Accuracy = 98.0%		Twin birth at the studied parity: No	Yes	Twin as 5th birth, b, first stage = 0.434, Accuracy = 98.1%		Twin birth at the studied parity: No	Yes
"Unwanted birth"	No	0.9725	0.0101	"Unwanted birth"	No	0.9729	0.0098
	Yes	0.0100	0.0073		Yes	0.0096	0.0077

The first-stage coefficient would be biased if there are unobserved differences between mothers who do and do not experience a twin birth. Bhalotra and Clarke (2016), for example, suggest that mothers who experience a twin birth are, on average, somewhat healthier than other mothers. This would increase the size of the first-stage coefficient if healthier mothers also have, on average, more children than less healthy mothers. Bhalotra and Clarke (2016, tab. 7–8), contrary to this argument, shows that the coefficient in the first-stage regression increases when they add controls for the health of the mothers. What their results do show is that unobserved differences between mothers who do or do not experience a (parity-specific) twin birth will affect also the first-stage regression coefficient.

7. The effect that is being estimated when using twin birth instrumental variables

That not all (parity-specific) twin births lead to a both unexpected and "*unwanted*", i.e. exogenous, increase in the number of children does not bias the results from the IV regression. What it does is to highlight the necessary assumptions discussed above. We also have to reduce the claims regarding estimating a generalizable, causal effect.

We don't carry out IV regressions because we are interested in if children are taller or shorter in families that experienced a twin birth at a specific parity or not. We want to be able to interpret the estimated effect as the causal effect from the number of children on the outcomes for the children. We get close to this but not all the way. What we estimate is the local average treatment effect; LATE (Angrist, Imbens, and Rubin 1996). The LATE effect is the effect on the outcome from increasing the number of children of parents "whose treatment status can be changed by the instrument" (e.g. Angrist and Pischke 2009). Angrist and Pischke (2009) argue that we should consider the treatment to be the "extra" child born because of the twin birth. But, the parents whose "treatment status", i.e. number of children, can be "changed by the instrument", i.e. a (parity-specific) twin birth, are the parents for whom the twin birth lead to an unintended increase of the number of children. To see why we have to use the counterfactual framework and consider what these parents would have done if they had not experienced a (parity-specific) twin birth. If the parents, for example, wanted two more children, they would (in the counterfactual absence of the twin birth) have continued to get another child. Their treatment status, again i.e. number of children, was therefore not "changed by the instrument", again i.e. the (parity-specific) twin birth. What we therefore estimate when we use twin births for instrumental variables is the local average treatment effect.

This is the effect on the outcome studied from an exogenous increase of the number of children in *some* families because of a (parity-specific) twin birth; the "Treatment Effect". It is the effect for this subpopulation rather than for all families or for all families experiencing a twin birth; the "Local" effect. The estimated effect is the average of the effect experienced by all families in this sub-population; the "Average" effect. When we use twin births as second births for an instrumental variable we estimate the effect from increasing the number of children to three in families that only intended to have two children. This is different from the effect of increasing the number of children from two to three in any family, but could still be of interest. It is less generalizable than what is currently sometimes assumed in the literature.

It is important that we only include families that have reached (or surpassed) their desired number of children in our sample when we use (parity-specific) twin births for instrumental variables. If we use (parity-specific) twin births for instrumental variables for families that have not yet reached (or surpassed) their desired number of children, this will lead to that the estimated effect from sibship size on the outcome is biased towards zero. As discussed above, if we study a population of parents that have not yet reached (or surpassed) their desired number of children we will overestimate the difference in number of children between families that do or do not experience a (parity-specific) twin birth, i.e. the first-stage regression coefficient is too large. The difference in the outcome will, in contrast, be a result only of the increase in the number of children in families who experienced an "unwanted" birth. This is what we assume in the exclusion restriction; that the older siblings are not (positively or negatively) affected by getting two younger siblings at once instead of with some time in-between. The bias toward zero is therefore a result of that we are overestimating the difference in sibship size but comparing this to the smaller, "true" effect on the outcome. The bias towards zero when we are studying families that have not yet reached (or surpassed) their desired number of children will therefore be present even if the instrument otherwise is valid.

There are several reasons for why we should expect the estimated effect to vary across versions of the twin birth instrumental variables and across populations. One reason is that the estimated effect is a (weighted) average marginal effect across all the levels of the endogenous variable that are indicated by the instrumental variable. When we use twin births for instrumental variables the estimator will place more weight to the effect at the sibship sizes for which the instrumental variable has the largest effect on the sibship size. Twin birth instrumental variables indicate families with a twin birth at the studied parity regardless of whether they have more children after that parity or not. The instrumental variables therefore indicate families with different number of children. While these larger families are also indicated by the instrumental variable, the earlier twin birth has had no effect on the number of children in these families. These are families that would have gone on to have more births if they had only had a single birth instead of the twin birth. The effect from sibship size on the outcome is therefore estimated from *only* the (exogenous) increase of the number of children *at the studied parity*. If the marginal effect from increasing the number of children is different at different parities, as suggested by Mogstad and Wiswall (2016), this will lead to different results when using instrumental variables based on twin births at different parities.

There is also a reason to expect the estimated effect to vary across versions of the twin birth instrumental variables and across populations that is caused by the heterogenous treatment effect. This is related to the varying accuracy of the classification made by the instrumental variable. The accuracy is the extent to which the classification of the instrumental variables indicates (one group of) “compliers”, i.e. families that wanted *n* children but had *n+1* because of the twin birth.

The accuracy varies across distributions of desired number of children, i.e. populations, and across versions of the instrumental variable. The parity-specific twin birth instrumental variables have a high accuracy in all populations. The reason for this is that almost all births are single births, and therefore in the simulation are, by construction, not “unwanted”. The variation between populations and versions of the instrumental variable comes from the other, much less common categories. When we compare these other categories we see that the groups that are being categorized wrongly is oftentimes at least as large as the group correctly classified. The accuracy is especially high in the lower fertility distribution populations (Table 2). But the accuracy varies between the birth orders, especially between the second and the later births. A smaller coefficient in the first-stage regression will coincide with a lower accuracy of the classification of families by the instrumental variable. One reason for why this should have us expect different estimated effects when we use different instrumental variables or study different populations is that the groups studied will be a differently select group in the different populations. Remember that we are estimating the effect of exogenously increasing the number of children in some of the families that wanted exactly *n* children. This will be a differently select group for different instrumental variables and populations (Table 1), thus making the estimated effect more or less generalizable in different applications.

8. Negative associations, but no negative effects in instrumental variables regressions

Researchers from many different fields are investigating the associations between sibship size and child outcomes, and especially the mechanisms behind the often-observed negative associations. The literature on this is expanding very rapidly at the moment. Much effort has been put into evaluating if the sibship size has a causal effect on the life chances of the children or not. Not the least economists have furthered this specific discussion. The results in most of these, and other, studies confirm the general finding of a negative association between the sibship size and the outcome studied. To cite a few well-known examples, Black, Devereux, and Salvanes (2005), Angrist, Lavy, and Schlosser (2010), as well as Åslund and Grönqvist (2010) do find that sibship size is negatively associated with education and labor market outcomes.

The important difference with these, and many other, studies is that they then move on to investigate if these negative associations is the result of a causal effect or not. There are, as mentioned, many reasons as to why the observed association might not be the actual, causal effect from the sibship size on child outcomes. The three studies just mentioned have in common that they use parity-specific twin births as instrumental variables (IVs) for the sibship size.

There has emerged a pattern in the results from these studies and many others like them. While these studies do find negative associations between sibship size and different child outcomes, they do not find any causal effect when they use twin births for instrumental variables to estimate this (again, for example, Black, Devereux, and Salvanes 2005; Angrist, Lavy, and Schlosser 2010; Åslund and Grönqvist 2010 in regard to child education and labor market outcomes). This pattern of finding negative associations but no causal effect using twin births for IVs has been reported also for other outcomes. Baranowska-Rataj, de Luna, and Ivarsson (2016), for example, find that having many siblings is associated with a reduced risk of using dermatological and respiratory systems medications, and an increased risk of using psycoanaleptics and psycholeptics, at age 45. These effects disappear when they use parity-specific twin births for IVs for the number of children.

This pattern of results, from a large number of studies, has called into question if there really is any causal effect from sibship size on child outcomes. But, there has recently been a number of studies that challenge the conclusions drawn from these IV regression studies. The well-known no-effect-results in Black, Devereux, and Salvanes (2005) are, for example, sensitive to the assumed linear functional form. When Mogstad and Wiswall (2016) estimate the effect from sibship size on education for the same population, but using a more flexible functional form they do find effects. Importantly they find that the effect is different for different sibship sizes, being positive in small families and negative in large. Because the estimates in Black, Devereux, and Salvanes (2005) are (weighted) averages across these different effects, they do not find any effect at all.

An even more serious challenge to the pattern of results—negative associations, but no effect when parity-specific twin births are used for IVs—comes from Bhalotra and Clarke (2016). They argue that instrumental variables based on twin births are invalid for many applications because they are only conditionally independent of the outcomes. They show convincingly, across a wide range of populations, that the likelihood of experiencing a twin birth is associated with a large number of usually not observed characteristics of the mother. The likelihood of a twin birth is positively associated with the mother's level of education (also in a sample not using artificial reproductive technologies), indicators of maternal health, health-related behaviors, access to health facilities, and height, and negatively associated with stress and malnutrition. If we use twin births to create instrumental variables without adjusting for these differences between mothers the instrumental variables will not be valid and the results will be biased. Bhalotra and Clarke (2016) also, importantly, show that the bias we should expect from the IVs not being valid could be part of the explanation for the pattern of results in previous research. When they apply the method proposed by Conley, Hansen and Rossi (2012) to adjust the results for the bias, this leads to the re-emergence of a statistically significant, negative effect from sibship size on child outcomes.

We should, of course, not disregard the results from all these studies using IV regressions. But, I think the issues raised in this chapter and by Mogstad and Wiswall (2016), and, especially, Bhalotra and Clarke (2016) are serious enough that we might have to reconsider them. Future research must continue to take the possible confoundedness of the association seriously and try different ways to address this. Using parity-specific twin births as instrumental variables for the sibship size has this far been considered the "gold standard" method for this line of research. The recent challenges to the method have highlighted the need to critically evaluate also "gold standard" methods. We should not be content with just one way of addressing problems with endogeneity.

Many instrumental variables applied in social sciences are "plausibly", but likely not strictly exogenous (Conley, Hansen and Rossi 2012). Unfortunately plausible exogeneity is not enough to get accurate results. Even mild violations of the assumptions required to make the instrument valid leads to biased results. We therefore always need to carefully consider if the instrumental variables we use are truly exogenous. The results in Bhalotra and Clarke (2016) force us to reevaluate the validity of using twin births for instrumental variables for many applications. I argue in this chapter that a seemingly slight change of how we think about instrumental variables based on twin births conceptually actually highlights some previously overlooked assumptions that further question their validity. At the very least this change in thinking shows that we need to reduce the claims of estimating generalizable, causal effects when using twin birth instrumental variables.

References

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin (1996). "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91 (434): 444–455. doi:10.2307/2291629.

Angrist, Joshua D., and Alan B. Krueger (2001). "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments." *Journal of Economic Perspectives* 15 (4): 69–85. doi:10.1257/jep.15.4.69.

Angrist, Joshua D., Victor Lavy, and Analia Schlosser (2010). "Multiple Experiments for the Causal Link between the Quantity and Quality of Children." *Journal of Labor Economics* 28 (4): 773–824. doi:10.1086/653830.

Angrist, Joshua David, and Jörn-Steffen. Pischke (2009). *Mostly Harmless Econometrics : An Empiricist's Companion*. Princeton: Princeton University Press.

Åslund, O., and H. Grönqvist (2010). "Family Size and Child Outcomes: Is There Really No Trade-Off?" *Labour Economics* 17 (1): 130–139. doi:10.1016/j.labeco.2009.05.003.

Baranowska-Rataj, Anna, Xavier de Luna, and Anneli Ivarsson (2016). "Does the Number of Siblings Affect Health in Midlife? Evidence from the Swedish Prescribed Drug Register." *Demographic Research* 35 (43): 1259–1302. doi:10.4054/DemRes.2016.35.43.

Becker, Gary S. (1993). *A Treatise on the Family*. Enlarged edition, Reprint. Cambridge, Mass.: Harvard University Press.

Bhalotra, Sonia, and Damian Clarke (2016). "The Twin Instrument." IZA DP No. 10405. http://ftp.iza.org/dp10405.pdf.

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes (2005). "The More the Merrier? The Effect of Family Size and Birth Order on Children's Education." *The Quarterly Journal of Economics* 120 (2): 669–700. doi:10.1093/qje/120.2.669

Blake, Judith (1981). "Family Size and the Quality of Children." *Demography* 18 (4): 421–442. doi:10.2307/2060941

Blake, Judith (1985). "Number of Siblings and Educational Mobility." *American Sociological Review* 50 (1): 84–94. doi:10.2307/2095342.

Blau, Peter M., and Otis Dudley Duncan (1978). *American Occupational Structure*. 1 edition. New York: Free Press.

Bollen, Kenneth A. (2012). "Instrumental Variables in Sociology and the Social Sciences." *Annual Review of Sociology* 38: 37–72. doi:10.1146/annurev-soc-081309-150141

Conley, Timothy G., Christian B. Hansen, and Peter E. Rossi (2012). "Plausibly Exogenous." *The Review of Economics and Statistics* 94 (1): 260–272. doi:10.1162/REST_a_00139.

Dribe, Martin, Marco Breschi, Alain Gagnon, Danielle Gauvreau, Heidi A. Hanson, Thomas N. Maloney, Stanislao Mazzoni, Joseph Molitoris, Lucia Pozzi, Ken R. Smith, and Hélène Vézina (2017). "Socio-Economic Status and Fertility Decline: Insights from Historical Transitions in Europe and North America." *Population Studies* 71 (1): 3–21. doi:10.1080/00324728.2016.1253857.

Gibbs, Benjamin G., Joseph Workman, and Douglas B. Downey (2016). "The (Conditional) Resource Dilution Model: State- and Community-Level Modifications." *Demography* 53 (3): 723–748. doi:10.1007/s13524-016-0471-0.

Kalmijn, Matthijs, and Herman G. van de Werfhorst (2016). "Sibship Size and Gendered Resource Dilution in Different Societal Contexts." PLOS ONE 11 (8): e0160953. doi:10.1371/journal.pone.0160953.

Mogstad, Magne, and Matthew Wiswall (2016). "Testing the Quantity–quality Model of Fertility: Estimation Using Unrestricted Family Size Models." *Quantitative Economics* 7 (1): 157–192. doi:10.3982/QE322

Murray, Michael P. (2006). "Avoiding Invalid Instruments and Coping with Weak Instruments." *Journal of Economic Perspectives* 20 (4): 111–132. doi:10.1257/089533006780387373

Öberg, Stefan (2015). "Sibship Size and Height Before, during and after the Fertility Decline." *Demographic Research* 32 (2): 29–74. doi:10.4054/DemRes.2015.32.2.

Park, Hyunjoon (2008). "Public Policy and the Effect of Sibship Size on Educational Achievement: A Comparative Study of 20 Countries." *Social Science Research* 37 (3): 874–887. doi:10.1016/j.ssresearch.2008.03.002.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Roberts, Evan, and John Robert Warren (*forthcoming* 2017). "Family Structure and Childhood Anthropometry in Saint Paul, Minnesota in 1918." *The History of the Family* 22 (2). doi:10.1080/1081602X.2016.1224729.

Rosenzweig, Mark R., and Kenneth I. Wolpin (1980). "Testing the Quantity-Quality Fertility Model: The Use of Twins as a Natural Experiment." *Econometrica* 48 (1): 227–240. doi:10.2307/1912026.

Rosenzweig, Mark R., and Kenneth I. Wolpin (2000). "Natural 'Natural Experiments' in Economics." *Journal of Economic Literature* 38 (4): 827–874. doi:10.1257/jel.38.4.827.

Silventoinen, Karri, Mikko Myrskylä, Per Tynelius, Yoshie Yokoyama, and Finn Rasmussen (2013). "Social Modifications of the Multiple Birth Effect on IQ and Body Size: A Population-Based Study of Young Adult Males." *Paediatric and Perinatal Epidemiology* 27 (4): 380–387. doi:10.1111/ppe.12054.

Staiger, Douglas, and James H. Stock (1997). "Instrumental Variables Regression with Weak Instruments." *Econometrica* 65 (3): 557–586. doi:10.2307/2171753.

Stock, James H., Jonathan H. Wright, and Motohiro Yogo (2002). "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments." *Journal of Business & Economic Statistics* 20 (4): 518–529. doi:10.1198/073500102288618658

Stradford, Laura, Frans van Poppel, and L. H. Lumey (*forthcoming* 2017). "Can Resource Dilution Explain Differences in Height by Birth Order and Family Size? A Study of 389,287 Male Recruits in Twentieth-Century Netherlands." *The History of the Family* 22 (2). doi:10.1080/1081602X.2016.1230510.

Xu, Jun (2008). "Sibship Size and Educational Achievement: The Role of Welfare Regimes Cross-Nationally." *Comparative Education Review* 52 (3): 413–436. doi:10.1086/588761.

Göteborg Papers in Economic History

Available online at S-WOPEC: (http://swopec.hhs.se/gunhis/)

1. Jan Bohlin: Tariff protection in Sweden 1885-1914. 2005
2. Svante Larsson: Globalisation, inequality and Swedish catch up in the late nineteenth century. Williamson's real wage comparisons under scrutiny. 2005
3. Staffan Granér: Thy Neighbour's Property. Communal property rights and institutional change in an iron producing forest district of Sweden 1630-1750. 2005
4. Klas Rönnbäck: Flexibility and protectionism. Swedish trade in sugar during the early modern era. 2006
5. Oskar Broberg: Verkstadsindustri i globaliseringens tidevarv. En studie av SKF och Volvo 1970-2000. 2006
6. Jan Bohlin: The income distributional consequences of agrarian tariffs in Sweden on the eve of World War I. 2006
7. Jan Bohlin and Svante Larsson: Protectionism, agricultural prices and relative factor incomes: Sweden's wage-rental ratio, 1877–1926. 2006
8. Jan Bohlin: Structural Change in the Swedish economy in the late nineteenth and early twentieth century – The role of import substitution and export demand. 2007
9. Per Hallén Levnadsstandarden speglad i bouppteckningar. En undersökning av två metoder att använda svenska bouppteckningar för en levnadsstandards undersökning samt en internationell jämförelse. 2007
10. Klas Rönnbäck: The price of sugar in Sweden. Data, source & methods. 2007
11. Klas Rönnbäck: From extreme luxury to everyday commodity – sugar in Sweden, 17th to 20th centuries. 2007
12. Martin Khan: A decisive intelligence failure? British intelligence on Soviet war potential and the 1939 Anglo-French-Soviet alliance that never was. 2008
13. Bengt Gärdfors: Bolagsrevisorn. En studie av revisionsverksamheten under sent 1800-tal och tidigt 1900-tal. Från frivillighet till lagreglering och professionalisering. 2010
14. Ann-Sofie Axelsson, Oskar Broberg och Gustav Sjöblom (red.): Internet, IT-boomen och reklambranschen under andra hälften av nittiotalet. Transkript av ett vittnesseminarium på ABF-huset i Stockholm den 17 februari 2010. 2011
15. Staffan Granér and Klas Rönnbäck: Economic Growth and Clean Water in the Göta River. A Pilot Study of Collective Action and the Environmental Kuznets Curve, 1895-2000. 2011
16. Ulf Olsson: En värdefull berättelse. Wallenbergarnas historieprojekt. 2013
17. Irene Elmerot: Skrivhandledning för doktorander i ekonomisk historia vid Göteborgs universitet.2015

18. Tobias Karlsson and Christer Lundh: The Gothenburg Population Panel 1915–1943. GOPP Version 6.0. 2015
19. Tobias Karlsson: Pushed into unemployment, pulled into retirement: Facing old age in Gothenburg, 1923–1943. 2015
20. Stefan Öberg and Klas Rönnbäck: Mortality among European settlers in pre-colonial West Africa: The ”White Man’s Grave” revisited. 2016
21. Dimitrios Theodoridis: The ecological footprint of early-modern commodities. Coefficients of land use per unit of product. 2017
22. Stefan Öberg: An introduction to using twin births as instrumental variables for sibship size. 2017