Statistical verification of a natural "natural experiment": Tests and sensitivity checks for the sibling sex ratio instrument

Martin Huber

University of St. Gallen, Dept. of Economics

Quant Sem

Oct 1st 2012

Outline of the talk

In a nutshell:

- Statistical verification of the sibling sex ratio instrument of Angrist and Evans (1998) who investigate the effect of fertility on female labor supply
  - Testing IV validity based on Kitagawa (2008) and Huber and Mellace (2011, 2012)
  - Sensitivity checks for deviations from IV validity
- Sibling sex ratio instument is found to be close to valid
- Negativity of the effect is not overthrown by the sensitivity checks

Outline:

- Introduction
- Identifying assumptions and testing
- Empirical application
- Sensitivity analysis
- Conclusion

Introduction

How to identify the labor supply effect of fertility:

- Fertility and female labor supply decisions are most likely endogenous.
- Therefore, the effect of fertility on female labor supply is not easily identified.
- Angrist and Evans (1998) suggest to use the sex ratio of the first two siblings as an instrument for fertility.
- Intuition: If parents have a preference for mixed sex children, having two children of the same sex, which is arguably randomly assigned by nature, increases the chances of getting a third child.
- They estimate the local average treatment effect on the compliers (those induced to get a third child if the first two have the same sex), relying on the IV exclusion restriction and the monotonicity of the treatment in the instrument (see Imbens and Angrist, 1994).

Is the IV exclusion restriction plausible?

- Rosenzweig and Wolpin (2000) argue that having mixed sex siblings may violate the exclusion restriction by directly affecting both the marginal utility of leisure and child rearing costs and, thus, labor supply.
- Based on data from rural India, they provide empirical evidence that expenses for clothing of the third born are significantly lower if the older siblings are of the same sex.
- It is unclear to which extent this issue carries over to the US data of Angrist and Evans (1998) and whether it affects female labor market behavior.
- Bütikofer (2010) finds no crucial differences in the household economies of scale across families with different sibling sex composition due to cloth- and room-sharing for the UK, Switzerland, and Mexico.

Is monotonicity plausible?

- Parental preferences for a balanced sex composition are well documented for the US, see Ben-Porath and Welch (1976).
- However, monotonicity fails even if only a small fraction of parents prefers two children of the same sex and chooses to have a third child if the first two are of mixed sex.
- E.g., Lee (2008) finds that South Korean parents with one son and one daughter are more likely to continue childbearing than parents with two sons.
- Dahl and Moretti (2008) find that the number of children is significantly higher in US families with a first-born girl suggesting a preference for having boys.
- This raises concerns about the monotonicity assumption, but the results do not provide a direct test for its violation.

Contribution (1):

- The main contribution of this paper is the statistical verification of the validity of the sibling sex ratio instrument, based on hypothesis tests proposed by Kitagawa (2008) and Huber and Mellace (2011, 2012).
- The tests are based on (i) the supremum of violations of IV validity over the potential outcome distribution of compliers, (ii) the sum of violations over the distribution, or (iii) violations in the compliers' mean potential outcomes.
- They are applied to the full sample as well as to subsamples conditional on covariates such as education, age, marital status, race, and the year of the first birth.
- Finally, they are applied after splitting the instrument to separately consider two first born sons and two first born daughters vs. mixed sex siblings (to tackle potential violations of monotonicity).

Contribution (2):

- As a second contribution, sensitivity checks in the presence of an invalid instrument are proposed which are tailored to the nonparametric LATE framework.
- They allow assessing the robustness of the results under a violation of (i) the exclusion restriction, (ii) the monotonicity assumption, and (iii) both assumptions together.
- This complements previously suggested methods which focus either exclusively on...
  - *deviations from the exclusion restriction:* Hirano, Imbens, Rubin, and Zhou (2000), Manski and Pepper (2000), Altonji, Elder, and Taber (2005), Hoogerheide and van Dijk (2006), Nevo and Rosen (2008), Flores and Flores-Lagunes (2010), Choi and Lee (2011), Conley, Hansen, and Rossi (2012), Kraay (2012), and Mealli and Pacini (2012).
  - *deviations from monotonicity:* Huber and Mellace (2010) and Richardson and Robins (2010).

Identifying assumptions and testing

Notation:

- Y: outcome of interest
- D: binary endogenous treatment
- Z: binary instrument
- $D(1)$ and $D(0)$: potential treatment states for $Z = 1, 0$
- $Y(1)$ and $Y(0)$: potential outcomes for $D = 1, 0$
- $D = Z \cdot D(1) + (1 - Z) \cdot D(0)$
- $Y = D \cdot Y(1) + (1 - D) \cdot Y(0)$

Types (according to D(1),D(0)):

Table 1: Types

Type T	D(1)	D(0)	Notion
a	1	1	Always takers
c	1	0	Compliers
d	0	1	Defiers
n	0	0	Never takers

Linking types to observed subgroups:

Table 2: Observed subgroups and types

Observed values of Z and D	Potential types T
$\{i : Z_i = 1, D_i = 1\}$	observation i belongs either to a or to c
$\{i : Z_i = 1, D_i = 0\}$	observation i belongs either to d or to n
$\{i : Z_i = 0, D_i = 1\}$	observation i belongs either to a or to d
$\{i : Z_i = 0, D_i = 0\}$	observation i belongs either to c or to n

Identifying assumptions and testing

Observed and unobserved densities:

- Let $f(y, D = d|Z = z)$ denote the (observed) joint density of the observed outcome and $D = d$ conditional on $Z = z$ for $d, z \in \{1, 0\}$.
- Let $f(y(d), T = t|Z = z)$ denote the unobserved joint density of the potential outcome and type t conditional on $Z = z$, where $t \in \{a, c, d, n\}$.

$$
\begin{aligned}
f(y, D = 1|Z = 1) &= f(y(1), T = c|Z = 1) + f(y(1), T = a|Z = 1), && (1)\\
f(y, D = 1|Z = 0) &= f(y(1), T = d|Z = 0) + f(y(1), T = a|Z = 0), && (2)\\
f(y, D = 0|Z = 1) &= f(y(0), T = d|Z = 1) + f(y(0), T = n|Z = 1), && (3)\\
f(y, D = 0|Z = 0) &= f(y(0), T = c|Z = 0) + f(y(0), T = n|Z = 0). && (4)
\end{aligned}
$$

Identifying assumptions and testing

The identifying assumptions (Imbens and Angrist, 1994):

Assumption 1:
$Z \perp (D(1), D(0), Y(1), Y(0))$ (joint independence),

- Assumption 1 implies the randomization of the instrument (such that it is unrelated with factors affecting the treatment and/or outcome) and the exclusion of direct effects on the outcome.
- Also the types, which are defined by the potential treatment states, are independent of the instrument.

Assumption 2:
$\Pr(D(1) \geq D(0)) = 1$ (monotonicity).

- Assumption 2 says that the potential treatment state of any individual does not decrease (positive monotonicity) in the instrument.
- The existence of defiers (type d) is ruled out because for the latter group, $D(1) < D(0)$.
- Monotonicity is implicitly assumed in the linear IV model (homogeneous first stage coefficient).

Identifying assumptions and testing

Implications for the densities:

- Under Assumption 1, $f(y(d), T = t|Z = 1) = f(y(d), T = t|Z = 0) = f(y(d), T = t)$, otherwise the potential treatment states and/or potential outcomes were not independent of the instrument.
- Under Assumption 2, $f(y(1), T = d)$ and $f(y(0), T = d)$ are equal to zero.

Therefore, equations (1) to (4) simplify to

$$f(y, D = 1|Z = 1) = f(y(1), T = c) + f(y(1), T = a), \quad (5)$$
$$f(y, D = 1|Z = 0) = f(y(1), T = a), \quad (6)$$
$$f(y, D = 0|Z = 1) = f(y(0), T = n), \quad (7)$$
$$f(y, D = 0|Z = 0) = f(y(0), T = c) + f(y(0), T = n). \quad (8)$$

By subtracting (6) from (5) and (7) from (8), the joint densities of the compliers under treatment and non-treatment are identified:

$$f(y, D = 1|Z = 1) - f(y, D = 1|Z = 0) = f(y(1), T = c), \quad (9)$$
$$f(y, D = 0|Z = 0) - f(y, D = 0|Z = 1) = f(y(0), T = c). \quad (10)$$

Identifying assumptions and testing

Identification and testable constraints:

- (9) and (10) are sufficient to show that the probability limit of the Wald estimator identifies the LATE on the compliers:

$$E[Y(1) - Y(0)|T = c] = \Delta_c = \frac{E(Y|Z = 1) - E(Y|Z = 0)}{E(D|Z = 1) - E(D|Z = 0)}. \tag{11}$$

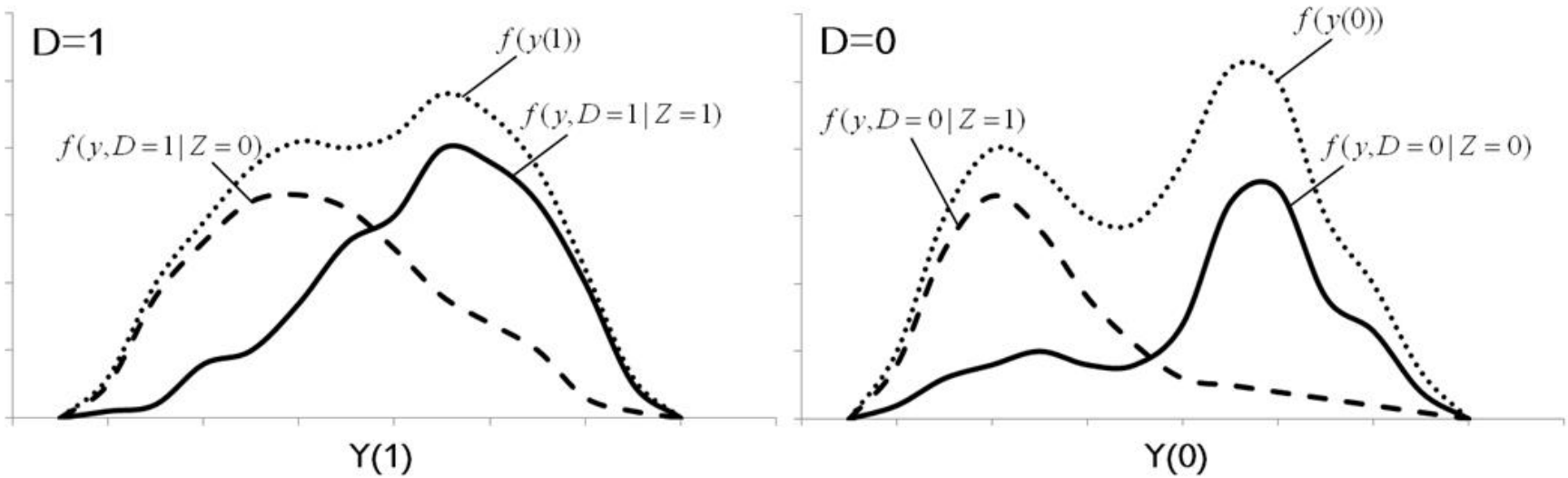
- (9) and (10) also have testable implications:

$$\begin{aligned} f(y, D = 1|Z = 1) &\geq f(y, D = 1|Z = 0), \\ f(y, D = 0|Z = 0) &\geq f(y, D = 0|Z = 1), \end{aligned} \tag{12}$$

- Under a violation of (12), the joint densities of the compliers would be negative, which is impossible.
- These constraints were first derived by Balke and Pearl (1997) for binary outcomes, while Kitagawa (2008) formulated them in terms of continuous outcomes.

Identifying assumptions and testing

Figure 1: Graphical illustration of the violations



Identifying assumptions and testing

Testing approaches:

(1) Kitagawa (2008) - testing probability constraints:

$$\begin{aligned} \Pr(Y \in A, D = 1|Z = 1) \geq \Pr(Y \in A, D = 1|Z = 0), \\ \Pr(Y \in A, D = 0|Z = 0) \geq \Pr(Y \in A, D = 0|Z = 1), \end{aligned} \tag{13}$$

where A denotes a subset of the support of Y.

(2) Huber and Mellace (2011) - testing (extended) prob. constraints:
In addition to (13), they consider the following constraints:

$$\begin{aligned} \Pr(Y \in A, D = 1|Z = 1) - \Pr(Y \in A, D = 1|Z = 0) \leq \Pr(D = 1|Z = 1) - \Pr(D = 1|Z = 0), \\ \Pr(Y \in A, D = 0|Z = 0) - \Pr(Y \in A, D = 0|Z = 1) \leq \Pr(D = 1|Z = 1) - \Pr(D = 1|Z = 0). \end{aligned} \tag{14}$$

- The intuition of (14) is that the joint probability of being a complier and having an outcome that lies in subset A cannot be larger than the (unconditional) complier share in the population.

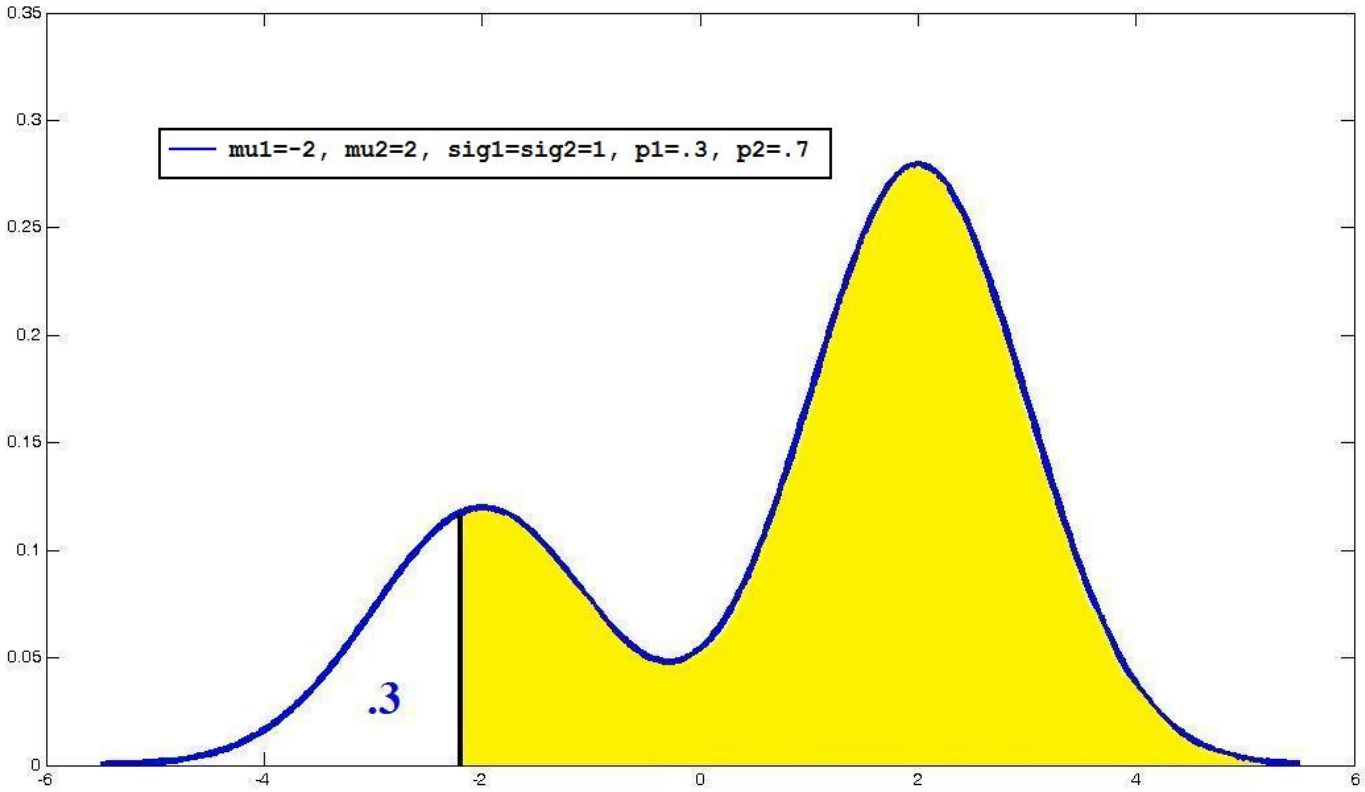
(3) Huber and Mellace (2011) - testing mean constraints:

$$E(Y|D=1,Z=1,Y\leq y_q)\leq E(Y|D=1,Z=0)\leq E(Y|D=1,Z=1,Y\geq y_{1-q}),$$
$$E(Y|D=0,Z=0,Y\leq y_r)\leq E(Y|D=0,Z=1)\leq E(Y|D=0,Z=0,Y\geq y_{1-r}), \quad (15)$$

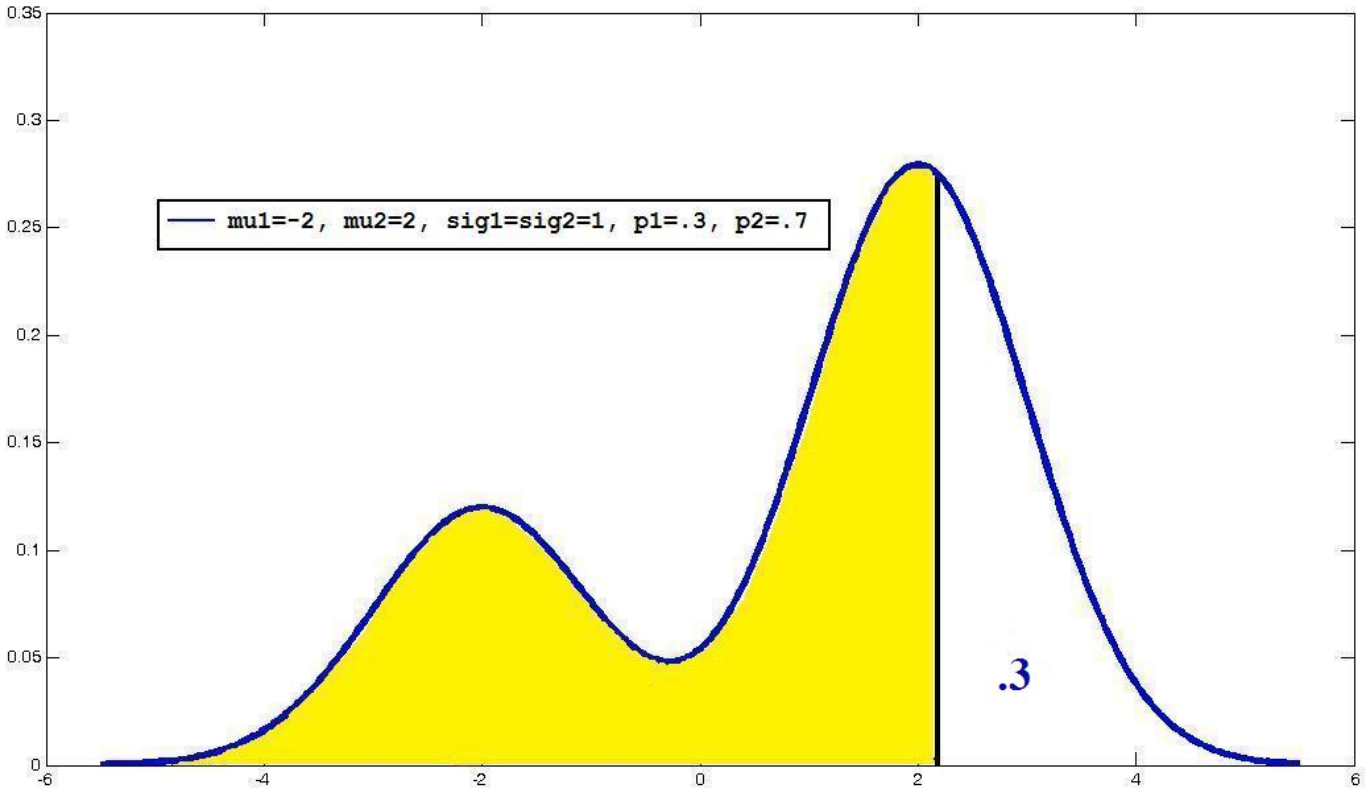
where $q=1-\frac{\Pr(D=1|Z=1)-\Pr(D=1|Z=0)}{\Pr(D=1|Z=1)}$ (the share of always takers among those with $D=1$ and $Z=1$) and $r=1-\frac{\Pr(D=1|Z=1)-\Pr(D=1|Z=0)}{\Pr(D=0|Z=0)}$ (the share of never takers among those with $D=0$ and $Z=0$).

- The point identified mean potential outcomes of the always takers under treatment and the never takers under non-treatment must not lie outside of their respective bounds in the mixed population (with the compliers).

Identifying assumptions and testing



Identifying assumptions and testing



(4) Huber and Mellace (2012) - testing integrals of densities:

$$\begin{aligned}
&\int [f(y, D=1|Z=0) - \min(f(y, D=1|Z=1), f(y, D=1|Z=0))]dy \\
= \quad &\Pr(D=1|Z=0) - \int \min(f(y, D=1|Z=1), f(y, D=1|Z=0))dy = 0, \\
&\int [f(y, D=0|Z=1) - \min(f(y, D=0|Z=0), f(y, D=0|Z=1))]dy \\
= \quad &\Pr(D=0|Z=1) - \int \min(f(y, D=0|Z=0), f(y, D=0|Z=1))dy = 0.
\end{aligned} \tag{16}$$

- If Assumptions 1 and 2 are satisfied,
  $\min(f(y, D=1|Z=1), f(y, D=1|Z=0)) = f(y, D=1|Z=0)$ because
  $f(y, D=1|Z=1) \geq f(y, D=1|Z=0)$ and
  $\min(f(y, D=0|Z=0), f(y, D=0|Z=1)) = f(y, D=0|Z=1)$ because
  $f(y, D=0|Z=0) \geq f(y, D=0|Z=1)$.
- Therefore, if the integrals are positive, the constraints in (12) are violated for at least one value of y in the support of Y.

Empirical application

The effect of fertility on female labor supply:

- Angrist and Evans (1998) estimate the effect of fertility on female labor supply based on the instrument sex composition of the two first children.
- Sample of 394,840 mothers who were aged between 21 and 35 years, had been 15 or older when giving birth for the first time, and had at least two children (with the second being at least one year old), coming from the 1980 wave of the U.S. Census Public Use Micro Samples (PUMS).

Empirical application

- D: indicator whether a mother has three or more children (158,751 observations or 40.21%) or just two kids (236,089 observations or 59.79%).
- Z: one if the first two children have the same sex (199,548 or 50.54%) and zero otherwise (195,292 observations or 49.46%).
- Y: hours worked per week (discrete).
- The first stage regression of D on Z is highly significant and suggests that if monotonicity holds, same sex increases the probability of a third child by roughly 6 percentage points.

Empirical application

Table 3: P-values when testing the sibling sex ratio instrument

		supremum tests		integral tests	
	Mean test	HM11 test	K08 test	for $D = 1$	for $D = 0$
full sample (394,840)	0.580	0.638	0.210	0.984	0.999
edu<12 (88,764)	1.000	0.778	0.745	0.869	0.890
edu=12 (189,818)	1.000	0.070	0.034	0.958	0.969
edu>12 (116,258)	1.000	0.961	0.996	0.980	0.999
edu<12, a. 20-25, married, white (11,716)	0.501	0.541	0.567	0.402	0.654
edu=12, a. 20-25, married, white (16,249)	0.520	0.092	0.048	0.892	0.196
edu>12, a. 20-25, married, white (3,232)	0.822	1.000	1.000	0.440	0.474
edu<12, a. 26-30, married, white (19,864)	0.617	0.030	0.022	0.922	0.597
edu=12, a. 26-30, married, white (53,919)	1.000	0.164	0.168	0.812	0.786
edu>12, a. 26-30, married, white (28,266)	1.000	0.304	0.376	0.848	0.812
edu<12, a. 31-35, married, white (23,367)	1.000	0.234	0.106	0.670	0.729
edu=12, a. 31-35, married, white (75,850)	1.000	0.368	0.378	0.947	0.994
edu>12, a. 31-35, married, white (58,684)	1.000	0.674	0.487	0.953	0.980
edu<12, a. 20-25, mar., w., 1st birth 70-72 (2,784)	0.604	0.426	0.559	0.170	0.551
edu=12, a. 20-25, mar., w., 1st birth 70-72 (1,713)	0.477	1.000	0.558	0.279	0.292
edu>12, a. 20-25, mar., w., 1st birth 70-72 (286)	0.153	1.000	0.061	0.331	0.031
edu<12, a. 26-30, mar., w., 1st birth 70-72 (8,592)	0.704	0.105	0.223	0.832	0.508
edu=12, a. 26-30, mar., w., 1st birth 70-72 (24,492)	1.000	0.277	0.228	0.610	0.619
edu>12, a. 26-30, mar., w., 1st birth 70-72 (9,149)	0.703	0.147	0.265	0.797	0.233
edu<12, a. 31-35, mar., w., 1st birth 70-72 (2,389)	0.624	1.000	0.536	0.348	0.224
edu=12, a. 31-35, mar., w., 1st birth 70-72 (15,913)	1.000	0.474	0.239	0.569	0.814
edu>12, a. 31-35, mar., w., 1st birth 70-72 (20,037)	1.000	0.502	0.403	0.868	0.976

Note: The p-values of the mean and supremum tests are based on 999 bootstrap draws. The p-values of the integral tests are based on 999 subsamples where the subsampling size corresponds to the integer closest to $n^{0.9}$.

Empirical application

Figure 2: Estimates of $f(y(1), T = c)$ and $f(y(0), T = c)$

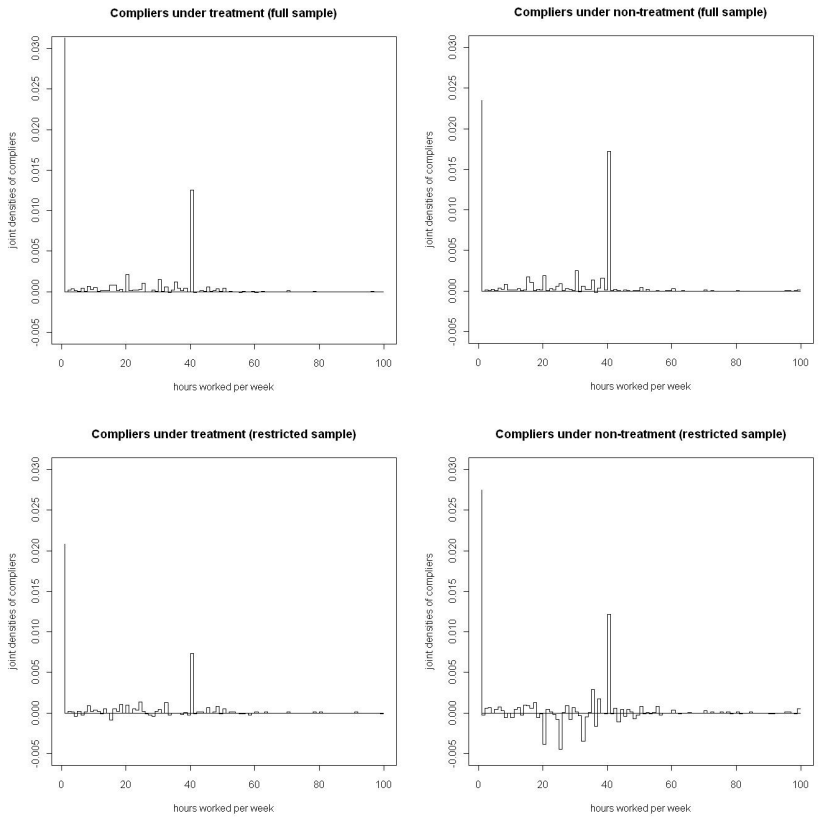


Figure 2 graphically shows the violations in the full sample (upper graphs) and conditional on age 21-25, 12 years of education, white, and married (lower graphs).

Empirical application

Table 4: Two girls instrument

		supremum tests		integral tests	
	Mean test	HM11 test	K08 test	for $D = 1$	for $D = 0$
full sample (290,713)	1.000	0.180	0.190	0.999	0.999
edu<12 (65,597)	1.000	0.888	1.000	0.973	0.829
edu=12 (139,674)	1.000	0.175	0.075	0.776	0.994
edu>12 (85,442)	1.000	0.194	0.264	0.969	0.979
edu<12, age 21-25, married, white (8,662)	0.515	0.057	0.427	0.370	0.699
edu=12, age 21-25, married, white (11,892)	0.709	0.086	0.012	0.410	0.203
edu>12, age 21-25, married, white (2,368)	0.638	1.000	0.903	0.495	0.238
edu<12, age 26-30, married, white (14,664)	1.000	0.032	0.024	0.931	0.713
edu=12, age 26-30, married, white (39,503)	1.000	0.184	0.149	0.784	0.967
edu>12, age 26-30, married, white (20,779)	1.000	0.453	0.505	0.807	0.874
edu<12, age 31-35, married, white (17,142)	1.000	0.471	0.381	0.853	0.828
edu=12, age 31-35, married, white (55,751)	1.000	0.529	0.461	0.882	0.992
edu>12, age 31-35, married, white (42,982)	1.000	0.191	0.358	0.953	0.786
edu<12, a. 21-25, mar., w., 1st birth 70-72 (2,064)	0.544	1.000	0.702	0.292	0.468
edu=12, a. 21-25, mar., w., 1st birth 70-72 (1,267)	0.576	1.000	0.510	0.156	0.133
edu>12, a. 21-25, mar., w., 1st birth 70-72 (204)	0.302	0.721	0.238	0.163	0.146
edu<12, a. 26-30, mar., w., 1st birth 70-72 (6,330)	0.703	0.256	0.205	0.455	0.573
edu=12, a. 26-30, mar., w., 1st birth 70-72 (17,903)	1.000	0.611	0.657	0.477	0.951
edu>12, a. 26-30, mar., w., 1st birth 70-72 (6,736)	1.000	0.492	0.452	0.507	0.238
edu<12, a. 31-35, mar., w., 1st birth 70-72 (1,762)	0.857	1.000	0.632	0.497	0.132
edu=12, a. 31-35, mar., w., 1st birth 70-72 (11,643)	0.920	0.910	1.000	0.711	0.875
edu>12, a. 31-35, mar., w., 1st birth 70-72 (14,690)	1.000	0.430	0.196	0.852	0.907

Note: The p-values of the mean and supremum tests are based on 999 bootstrap draws. The p-values of the integral tests are based on 999 subsamples where the subsampling size corresponds to the integer closest to $n^{0.9}$.

Empirical application

Table 5: Two boys instrument

		supremum tests		integral tests	
	Mean test	HM11 test	K08 test	for $D = 1$	for $D = 0$
full sample (299,419)	1.000	0.502	0.452	0.963	0.985
edu<12 (67,183)	1.000	0.594	0.485	0.640	0.919
edu=12 (143,863)	1.000	0.094	0.057	0.987	0.812
edu>12 (88,373)	1.000	0.681	0.528	0.918	0.985
edu<12, age 21-25, married, white (8,835)	0.522	0.540	0.493	0.309	0.511
edu=12, age 21-25, married, white (12,283)	0.495	0.513	0.283	0.920	0.247
edu>12, age 21-25, married, white (2,442)	0.677	1.000	0.360	0.211	0.679
edu<12, age 26-30, married, white (15,072)	0.650	0.063	0.327	0.823	0.239
edu=12, age 26-30, married, white (40,946)	1.000	0.338	0.300	0.832	0.423
edu>12, age 26-30, married, white (21,543)	0.511	0.352	0.474	0.621	0.766
edu<12, age 31-35, married, white (17,882)	0.521	0.089	0.096	0.409	0.559
edu=12, age 31-35, married, white (57,592)	1.000	0.536	0.490	0.978	0.991
edu>12, age 31-35, married, white (44,665)	1.000	0.182	0.085	0.828	0.967
edu<12, a. 21-25, mar., w., 1st birth 70-72 (2,068)	0.704	0.595	0.519	0.083	0.638
edu=12, a. 21-25, mar., w., 1st birth 70-72 (1,294)	0.577	1.000	0.756	0.371	0.306
edu>12, a. 21-25, mar., w., 1st birth 70-72 (226)	0.122	1.000	0.056	0.151	0.014
edu<12, a. 26-30, mar., w., 1st birth 70-72 (6,491)	0.653	0.186	0.589	0.575	0.239
edu=12, a. 26-30, mar., w., 1st birth 70-72 (18,685)	1.000	0.348	0.323	0.880	0.576
edu>12, a. 26-30, mar., w., 1st birth 70-72 (7,022)	0.687	0.148	0.225	0.310	0.420
edu<12, a. 31-35, mar., w., 1st birth 70-72 (1,817)	0.580	1.000	0.300	0.071	0.499
edu=12, a. 31-35, mar., w., 1st birth 70-72 (12,076)	1.000	0.232	0.132	0.380	0.723
edu>12, a. 31-35, mar., w., 1st birth 70-72 (15,265)	0.617	0.103	0.322	0.679	0.989

Note: The p-values of the mean and supremum tests are based on 999 bootstrap draws. The p-values of the integral tests are based on 999 subsamples where the subsampling size corresponds to the integer closest to $n^{0.9}$.

Violation of the exclusion restriction:

The LATE expression has to be adjusted for the direct effect (γ) :

$$\Delta_c = \frac{E(Y|Z=1) - \gamma - E(Y|Z=0)}{E(D|Z=1) - E(D|Z=0)}. \quad (17)$$

This corresponds to the probability limit of the slope coefficient in TSLS when regressing D on one and Z in the first stage and $Y - Z\gamma$ on one and the first stage prediction of D in the second stage.

Violation of monotonicity:

Note that under Assumptions 1 and 2,

$$E[Y(1)|T=c] = \frac{\Pr(D=1|Z=1)\cdot E(Y|D=1,Z=1) - \Pr(D=1|Z=0)\cdot E(Y|D=1,Z=0)}{\Pr(D=1|Z=1) - \Pr(D=1|Z=0)},$$
$$E[Y(0)|T=c] = \frac{\Pr(D=0|Z=0)\cdot E(Y|D=0,Z=0) - \Pr(D=0|Z=1)\cdot E(Y|D=0,Z=1)}{\Pr(D=1|Z=1) - \Pr(D=1|Z=0)}, \quad (18)$$

If monotonicity does not hold, it can be shown that

$$E[Y(1)|T=c] = \frac{\Pr(D=1|Z=1)\cdot E(Y|D=1,Z=1)}{\rho_a(\Pr(D=1|Z=0) - \pi_d) + \Pr(D=1|Z=1) - \Pr(D=1|Z=0) + \pi_d},$$
$$E[Y(0)|T=c] = \frac{\Pr(D=0|Z=0)\cdot E(Y|D=0,Z=0)}{\rho_n(\Pr(D=0|Z=1) - \pi_d) + \Pr(D=1|Z=1) - \Pr(D=1|Z=0) + \pi_d}, \quad (19)$$

where $\rho_a = E[Y(1)|T=a]/E[Y(1)|T=c]$, $\rho_n = E[Y(0)|T=n]/E[Y(0)|T=c]$, and $\pi_d = \Pr(T=d)$ (the share of defiers in the population).

Sensitivity analysis

How to choose the sensitivity parameters γ, ρ_a, ρ_n, and π_d?

- Concerning ρ_a and ρ_n, one may consider their empirical values in the data (if Assumptions 1 and 2 were satisfied) and then investigate plausible deviations from IV validity.
- In our sample, the estimated $\rho_a = 1.11$ and the estimated $\rho_n = 1.02$.
- An inspection of the pre-treatment covariates of the various populations may help to determine plausible deviations from these values.
- A caveat is that the covariate values may be affected by a violation of monotonicity (in contrast, the exclusion restriction should hold).

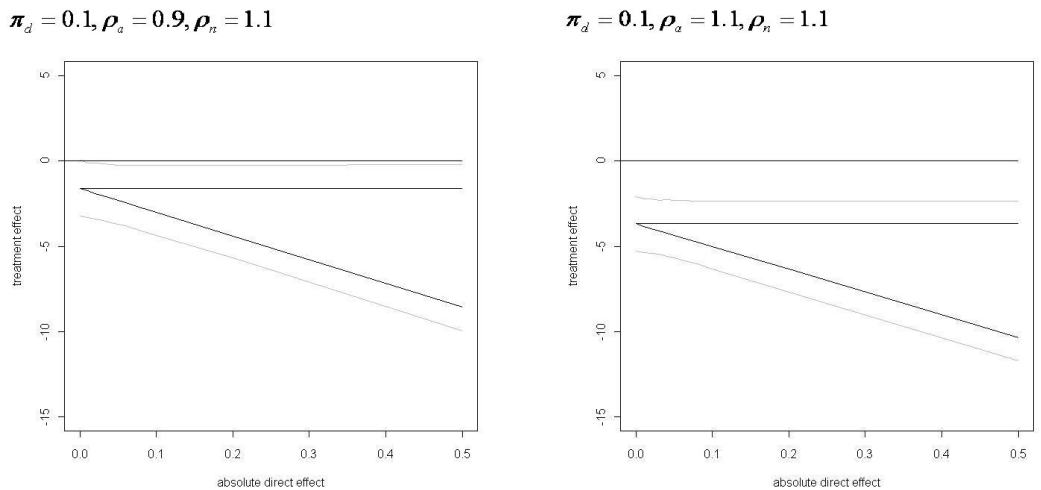
Table 6: Differences in observed covariates

	D=1			D=0			
Variable	always takers	compliers	diff	never takers	compliers	diff	LATE
Education	11.646	12.041	-0.395**	12.446	12.093	0.353**	-0.052
Age	30.559	30.346	0.213**	29.804	30.618	-0.815**	-0.272
Marital status	0.848	0.897	-0.048**	0.850	0.887	-0.037**	0.010

Note: **: Significant at the 1% level. *: Significant at the 5% level. +: Significant at the 10% level.

Sensitivity analysis

Figure 3: Bounds on the LATE (black lines) and 95% confidence intervals (grey lines)



- γ assumed to be nonnegative: increased costs of mixed sex siblings (no room-sharing and hand-me-downs) should (if anything) induce women to provide more labor to increase household income.
- ρ_a is varied btw. 0.9 and 1.1, as labor market relevant characteristics of always takers and compliers are not too different.
- ρ_n is set to 1.1, implying higher labor market attachment of never takers given the slightly more favorable characteristics and their choice not to have further children.
- π_d determines the width of the bounds, not the direction of the effect.

Conclusion

- Statistical verification of the sib sex ratio instrument of Angrist and Evans (1998) based on hypothesis tests
- Sensitivity analysis allowing for both violations of the exclusion restriction and monotonicity
- The tests do not point to important violations of the IV assumptions
- The negativity of the effect of fertility on female labor supply is robust to moderate violations of the IV assumptions