Instrumental variables

Clément de Chaisemartin

Majeure Economie

September 2011

1. Intuition
   - Why do we need instrumental variables ?
   - Definition and examples of instruments
2. Instrumental variables
   - Assumptions in the one variable / one instrument case
   - Properties in the one variable one instrument case
   - Are assumptions testable ?
   - Limits of instrumental variables
   - Back to intuition
3. Local average treatment effect
   - IV in randomized experiments
   - The Rubin framework of potential outcomes
4. Applications
   - Returns to education
   - Impact of the number of children on women labor supply
5. Conclusion

Issues with OLS (1/2)

- We assume we run the following regression: $Y = \alpha + \beta_1 X_1 + \varepsilon$. But the assumption that $cov(X_1, \varepsilon) = 0$ does not hold. Potential reason: the true data generating process is $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \varepsilon'$. Omitted variable bias.
- Wage and education example: education correlated to cleverness.
- In this case, $\widehat{\beta_1}$ is not asymptotically consistent: the data does not allow telling to what extent the second story is true.
- Omitted variable bias will happen all the time. Particularly strong in the wage / education regression due to ability bias (ability impossible to observe statistically).
- Impossible to run a randomized experiment on this issue: you can not compell the control group to leave school at age 14 and you can not oblige the test group to go for a PhD.

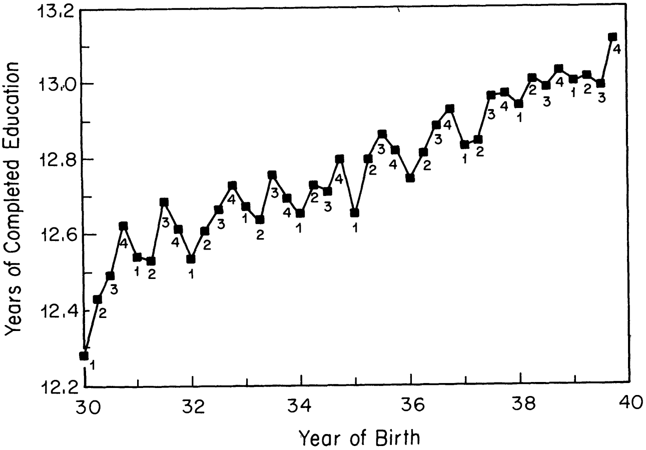
Two conditions to define an instrument (1/2)

- If Z is correlated to X_1 and Z is not correlated to Y once the effect of X_1 on Y is controlled for (that is to say that Z is not correlated to ε), then Z can be used as an instrumental variable to measure the effect of X_1 on Y.
- Example: the fact to be born in the first quarter of the year has an impact on the number of years of education you will follow if you are an American citizen.
    - schooling becomes compulsory in the beginning of the schooling year of the civil year during which you will turn / have turned 6 years old. => you enter school at 5 years and 9 months if you were born in December, 6 years and 8 months if January. End of mandatory schooling = when you turn 16.
    - Z (a variable for quarter of birth) is likely to be correlated with X_1, the number of years of education completed by an individual.

Two conditions to define an instrument (2/2)

- Once this small effect of Z on X_1 is accounted for, there is just no reason why Z should have any impact on Y, that is to say on your income.
- Quarter of birth (QOB) can therefore be used as an instrument to measure the impact of the number of years of education on wages. Paper by Angrist and Krueger.

Some evidence that QOB influences education



Reduced form (1/2)

- A first thing to be done would be to merely regress wages just on QOB, or on a dummy variable for individuals born in the first quarter of the year. This you know how to do it.
- $Y = \alpha + \beta 1_{\{born\ first\ quarter\}} + \varepsilon$.
- If you find that $\widehat{\beta} = -100$ and is statistically significant, then under which assumption can you conclude that education has an impact on wages ?
- However, would this regression give you some information on the size of the effect of education on wages ? Can you claim for instance that 1 supplementary year of education increases income by 500€ ? What is missing to make such a claim ?

Reduced form (2/2)

- Under the assumption that the fact of being born in the first quarter of the year has an impact on wages only through its impact on education, then the fact that $\widehat{\beta} = -100$ and is statistically significant means that education has an impact on wages.
- You know the relationship between Y and $1_{\{born\ first\ quarter\}}$. To be able to translate this into a relationship between Y and *education* you need to know what is the relationship between *education* and $1_{\{born\ first\ quarter\}}$:
  - Assume that the fact of being born in the first quarter of the year has a very large impact on education: reduces educational attainment by 2 years.
  - => then one suplementary year of education increases wage by $\frac{-100}{-2} = 50€$. To find the effect of X on Y, you measure the effect of Z on Y and then you rescale it by the effect of Z on X.
  - On the contrary, if Z has a very small impact on X: reduces educational attainment by 0.1 year, then the 100€ difference in wages of people born in the first quarter and others means that one more year of education has a large impact on wages: $\frac{-100}{-0.1} = 1000€$.

Evidence that QOB has an impact on wages

Panel B: Wald Estimates for 1980 Census—Men Born 1930–1939

	(1) Born in 1st quarter of year	(2) Born in 2nd, 3rd, or 4th quarter of year	(3) Difference (std. error) (1) − (2)
ln (wkly. wage)	5.8916	5.9027	−0.01110
			(0.00274)
Education	12.6881	12.7969	−0.1088
			(0.0132)

Assumptions

- Assume that Y is generated according to the following data generating process: $Y = \alpha + \beta X + \varepsilon$ with $cov(X, \varepsilon) \neq 0$.
- Assume there exists a variable Z which fulfills the two following conditions:
  - $cov(Z, X) \neq 0$: Z is related to X
  - $cov(Z, \varepsilon) = 0$: Z is not related to all the other determinants of Y.
  - Putting it in other words, Z is related to Y because Z explains X and X explains Y. But Z is related to Y only through X: once the impact of X on Y is accounted for, Z has no longer any impact on Y.
  - if Z fulfills these two conditions, then it is called an instrumental variable.
- You can re-read this slide replacing Y by "wages", X by "number of years of education" and Z by "QOB".

Identification

- Under these two assumptions, $cov(Z,Y) = cov(Z,\alpha+\beta X+\varepsilon) = \beta cov(Z,X) \Rightarrow \beta = \frac{cov(Z,Y)}{cov(Z,X)}$.
- What is a natural estimator $\widehat{\beta}$ of β according to this formula ?

Estimation: sample equivalents based on identification

- Under these two assumptions, $cov(Z,Y) = cov(Z, \alpha + \beta X + \varepsilon) = \beta cov(Z,X) \Rightarrow \beta = \frac{cov(Z,Y)}{cov(Z,X)}$.
- $\widehat{\beta}_{IV} = \frac{cov_e(Z,Y)}{cov_e(Z,X)}$.
- If there was no endogeneity issue, we could directly use the OLS and we would have $\widehat{\beta}_{OLS} = \frac{cov_e(X,Y)}{cov_e(X,X)}$. What can you notice ?
- Is the $\widehat{\beta}_{IV}$ estimator consistent ?

Estimation: double least squares (1/2)

- Assumption 1: $Y = \alpha + \beta X + \varepsilon$ with $cov(X, \varepsilon) \neq 0$.
- Assumption 2: $X = \gamma + \delta Z + \varepsilon^{'}$ with $cov(Z, \varepsilon) = 0$ which implies that $cov(\varepsilon^{'}, \varepsilon) \neq 0$.
- => X = the sum of two things: one thing which is uncorrelated with ε $(\gamma + \delta Z)$ and one thing which is correlated with ε $(\varepsilon^{'})$.
- => intuition = instead of regressing Y on X, regress Y only on the part of X which is uncorrelated to ε. Instead of X, we would like to keep only $X^* = \gamma + \delta Z$. Instead of comparing wages of individuals with different levels of education, compare wages of individuals who have different levels of education only because they were born in different quarters.

Estimation: double least squares (2/2)

- Issue: we do not observe δ and γ. But we can get good estimates of them with the OLS. => two stages procedure:
- Regress X on Z. Compute $\widehat{X} = \widehat{\gamma} + \widehat{\delta}Z$. First stage.
  - Regress Y on $\widehat{X}$. Second stage.

Assumption 1

- The assumption that $cov(Z, X) \neq 0$ can easily be tested.
  - Run the OLS regression $X = \gamma + \delta Z + \varepsilon'$.
  - If $\widehat{\delta}$ is significantly different from 0 (measured through a t-test) then there is some correlation between X and Z.

Assumption 2

- It is not possible to test the assumption that $cov(\varepsilon, Z) = 0$.
  - We never observe residuals of a regression. But we can observe estimates : $\widehat{\varepsilon} = Y - \widehat{Y}$. A test of whether $cov(\varepsilon, Z) = 0$ would be to compute $cov_e(\widehat{\varepsilon}, Z)$ and to see if significantly different from 0.
  - But whatever the instrument Z we will always have by construction $cov_e(\widehat{\varepsilon}, Z) = 0$. Proof:

$$cov_e(\widehat{\varepsilon}, Z) = cov_e(Y_i - \widehat{\alpha} - \frac{cov_e(Z_i, Y_i)}{cov_e(Z_i, X_i)}X_i, Z_i)$$

$$= cov_e(Y_i, Z_i) - cov_e(\widehat{\alpha}, Z_i) - \frac{cov_e(Z_i, Y_i)}{cov_e(Z_i, X_i)}cov(X_i, Z_i) = 0$$

  whether Z is truly uncorrelated with ε or not. Intuition: in the second stage, we regress Y on $\widehat{X} = \widehat{\gamma} + \widehat{\delta}Z$=> Z is an explanatory variable in this regression => by FOC2 of OLS $cov_e(\widehat{\varepsilon}, Z) = 0$.

- No formal test => is it a "credible" assumption ? You must find "credible" stories (as with the QOB example) which could explain why the assumption is not valid.

What happens when assumptions are not verified...

- If $cov(\varepsilon, Z) \neq 0$, $\beta = \frac{cov(Z,Y)}{cov(Z,X)} - \frac{cov(Z,\varepsilon)}{cov(Z,X)}$. Since $\widehat{\beta}_{IV} \rightarrow \frac{cov(Z,Y)}{cov(Z,X)}$, $\widehat{\beta}_{IV} \rightarrow \beta + \frac{cov(Z,\varepsilon)}{cov(Z,X)}$. The estimator is no longer consistent: asymptotically biased.
- Even if the correlation between Z and ε is small, if the correlation between Z and X is small, then the asymptotic bias can be large. Weak instrument issue.
- When endogeneity issue, you remember that $\widehat{\beta}_{OLS} \rightarrow \beta + \frac{cov(X,\varepsilon)}{V(X)}$. => if $|\frac{cov(X,\varepsilon)}{V(X)}| < |\frac{cov(Z,\varepsilon)}{cov(Z,X)}|$, it is even worse to use IV than OLS.

Weak instruments

- => one should avoid using weak instruments:
- when assumptions are not verified, the asymptotic bias can be very large with a weak instrument.
  - rule of thumb: when regressing X on Z, the t-statistic should be greater than 3.5.
- Do you think that QOB is a weak instrument ?

Intuition (1/2)

- Let us get back to the QOB example.
- Some individuals have a higher level of education than others because:
  - they come from richer backgrounds => their parents can pay for the university
  - they come from more educated families => their parents incentivize them to study because they know it is important
  - they have a higher level of ability => they get a merit scholarship which allows them to go to university.
- All these variations in educational level are endogenous to wage: they are determined by variables which also have an influence on wage (parents salary, parents level of education, ability...)

Intuition (2/2)

- But some individuals get less education than others just because they were born in Q1:
  - there is no reason why being born in Q1 should be influenced by parents wages, parents educational level...
- This variation in educational level is exogenous to wages: not determined by another factor which is also a determinant of wages.
- => IV computes a new version of X netted out of all its other determinants than the instruments and sees how this new version of X relates to Y.
- We create a new version of educational level netted out of the effect of parents wage, parents educational level etc..., influenced only by QOB.

A question...

- Intuitively, do you think that in the QOB example, $\widehat{\beta}_{IV}$ measures the impact on wage of having a masters instead of having only a bachelor degree ?

Randomized experiments with imperfect compliance (1/2)

- Assume you want to measure the impact of a training program on unemployed but you do not want to (cannot) prevent unemployed from the control group to attend the training and you can not oblige those in the test group to attend it. Compliance issue.
- What you can do = randomly sort out one test group and one control group, send letters to the test group informing them that a training program will take place and do not inform the others. Encouragement design.
- $Z = 1$ if a letter is sent, $X = 1$ if the person followed the training program, $Y = 1$ if she had found a job after 6 months.

Randomized experiments with imperfect compliance (2/2)

- You can assume that X is not independent from Z. More people will follow the training among those who were sent a letter than among the others.
- Once the effect of X on Y is accounted for, it is natural to assume that Z is not correlated to Y outside its impact on X:
  - by definition, Z is uncorrelated to all the other determinants of Y: because Z was determined by a lottery, people who received the letter = similar to those who did not receive it with respect to age, gender, education...
  - Z will not have any impact per se on Y: receiving a letter cannot have an impact per se on you finding a job.

=> Thanks to randomization, Z is a valid instrument to estimate the effect of X on Y.

Interpreting the IV coefficient in the Rubin framework

- Here, Z, Y and X are binary. One can show than in such instances $\widehat{\beta_{IV}} = \frac{P(\widehat{Y=1|Z}=1)-P(\widehat{Y=1|Z}=0)}{P(\widehat{X=1|Z}=1)-P(\widehat{X=1|Z}=0)}$
- According to this formula, the effect of the program on the probability to find a job is measured as follows:

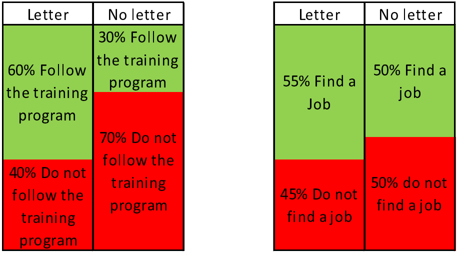
$$\frac{\%\ of\ \ letter\ group\ which\ found\ a\ job - \%\ of\ \ no\ letter\ group\ which\ found\ a\ job}{\%\ of\ \ letter\ group\ which\ followed\ training - \%\ of\ \ no\ letter\ group\ which\ followed\ training}.$$

- Does this quantity mean something inside the Rubin framework of potential outcomes (no longer assumption that treatment effect is constant) ?

Always takers, never takers and compliers

- Let us place ourselves in the Rubin framework of potential outcomes. Every individual has two potential outcomes regarding the fact of finding a job: $Y(1)$ if he follows the training, $Y(0)$ is he does not.
- We assume there are three types of people in our sample:
    - always takers: whether they receive the letter or not, they will participate to the training program.
    - never takers: whether they receive the letter or not, they will not participate to the training program.
    - compliers: they participate to the program if and only if they receive the letter.
- What we are assuming is that there are no people who participate to the program if they do not receive the letter and will not participate to it if they receive the letter (defiers).

Graphical representation



% of always takers, never takers and compliers

- Since we are observing only one scenario for each individual, we do not know who is a complier, who is an always taker etc...
- But we know:
    - the % of always takers: % of the no letter group which followed the training
    - the % of never takers: % of the letter group which did not follow the training
    - the % of compliers: % of the letter group which followed the training (compliers + always takers) - % of the no letter group which followed the training (always takers).
    - Why can we be sure that the percentage of always takers = the same in the test and in the control group ?

Understanding the IV formula for binary variables (1/2)

- Two groups of people. Some change their behavior due to the letter (compliers), some do not (always and never takers).
- This last group, whether they receive or not the letter, they do the same thing (always takers follow the training program, never takers do not follow it) => their employment status after 6 months cannot be affected by the letter.
- => the difference
  % with a job after 6 months in the letter group - % with a job after 6 months in the no letter group
  cannot come from never takers or always takers.
- => It can come only come from compliers.
- => since what we are measuring is only the effect of the training program on compliers, if we just measure it by
  % of the test group which found a job −
  % of the control group which found a job we are underestimating the true effect since not all individuals in our sample are compliers => we must divide this effect by the % of compliers in our sample. Kind of a rescaling.

Understanding the IV formula for binary variables (2/2)

- If 90% of the population are compliers, and we observe a small change in the probability to find a job (+ 2% from the no-letter to the letter group), then it means that the effect of the program is small.
- Most of the population changes its behavior when a letter is sent (attended the program whereas they would not have without the letter) and still, no big change in the probability of finding a job. Small effect on a large group.
- Whereas if only 5% of the population are compliers: $\%$ *of the test group which followed the training* $-$ $\%$ *of the control group which followed the training* $= 5\%$, then the 2% change we observe from the no-letter to the letter group comes only from 5% of the population => big effect on a small share of the population.

IV measure a local average treatment effect

- Under the assumption that there are no defiers (no people who will not attend the training program because of the letter), you can measure a causal effect through randomized experiments even when compliance is imperfect.
- However, this is not the ATT on the whole population that is measured, only the ATE on a given part of the population: the compliers. Hence the name local average treatment effect (LATE).
- To measure the LATE, run an instrumental variable regression of Y on X, using Z, the result of the lottery, as an instrument for X. If you regress Y directly on X, biased. Why ?
- In the QOB example, what is measured is the impact of being more educated on those who continued school because they were born in Q2, Q3 or Q4 and who would have stopped otherwise. => very specific population: the majority of the population will continue school after 16 whatever its quarter of birth.

Angrist and Krueger: returns to education

	(1) Born in 1st quarter of year	(2) Born in 2nd, 3rd, or 4th quarter of year	(3) Difference (std. error) (1) − (2)
ln (wkly. wage)	5.8916	5.9027	−0.01110
			(0.00274)
Education	12.6881	12.7969	−0.1088
			(0.0132)
Wald est. of return to education			0.1020
			(0.0239)
OLS return to education			0.0709
			(0.0003)

Comments on Angrist and Krueger

- Interpretation of the coefficients ?
- Is the IV estimate very different from the OLS ?
- How is the IV standard error wrt to the OLS ?
- Do you think their result properly measures the impact of having a masters degree instead of a bachelor ?
- Do you know many assets which bring you 10% per year ?
- Their result was criticized due to the weak instrument issue.

Children and women labor supply

- In aging countries, higher fertility and higher rate of participation of women to the labor market are two desirable objectives (financing of pensions, greater GDP...)
- We observe that women who participate to labor market have less children. But is it because working prevents them from having children (causal effect) or is it a mere correlation: women who work have less kids but this is determined by a third variable (taste for labor negatively related to taste for having a large family for instance, women who work are more educated => more able to control their fertility) ?
- Important policy question:
  - Germany launched a big investment program recently to increase the number of kindergarten...
  - If causal relationship, all type of facilities which make life easier for working mothers should be subsidized because this can raise birth rate along with labor market participation. If mere correlation, these expenses will just be a waste of money, will not incentivize mothers to work or working women to be mothers.

Angrist and Evans: Same sex children

- Idea of the paper= couples whose two oldest kids have the same sex are more likely to have a third one: preference for variety.
- Among couples with at least two children, the fact to have two kids with the same sex = random.
- => the variable: having two kids with the same sex = a good instrument to measure the impact of the number of kids on labor market participation.

Are IV assumptions verified ? (1/2)

Sex of first two children in families with two or more children	All women				Married women			
	1980 PUMS (394,835 observations)		1990 PUMS (380,007 observations)		1980 PUMS (254,654 observations)		1990 PUMS (301,588 observations)	
	Fraction of sample	Fraction that had another child	Fraction of sample	Fraction that had another child	Fraction of sample	Fraction that had another child	Fraction of sample	Fraction that had another child
one boy, one girl	0.494	0.372 (0.001)	0.495	0.344 (0.001)	0.494	0.346 (0.001)	0.497	0.331 (0.001)
two girls	0.242	0.441 (0.002)	0.241	0.412 (0.002)	0.239	0.425 (0.002)	0.239	0.408 (0.002)
two boys	0.264	0.423 (0.002)	0.264	0.401 (0.002)	0.266	0.404 (0.002)	0.264	0.396 (0.002)
(1) one boy, one girl	0.494	0.372 (0.001)	0.495	0.344 (0.001)	0.494	0.346 (0.001)	0.497	0.331 (0.001)
(2) both same sex	0.506	0.432 (0.001)	0.505	0.407 (0.001)	0.506	0.414 (0.001)	0.503	0.401 (0.001)
difference (2) − (1)	—	0.060 (0.002)	—	0.063 (0.002)	—	0.068 (0.002)	—	0.070 (0.002)

Notes: The samples are the same as in Table 2. Standard errors are reported in parentheses.

Are IV assumptions verified ? (2/2)

- Is the fact of having two first children with the same sex a random phenomenon ?
- Does it have an influence on the fact of having a third child ?

Wald estimator results

TABLE 5—WALD ESTIMATES OF LABOR-SUPPLY MODELS

	1980 PUMS			1990 PUMS			1980 PUMS		
		Wald estimate using as covariate:			Wald estimate using as covariate:			Wald estimate using as covariate:	
Variable	Mean difference by *Same sex*	*More than 2 children*	*Number of children*	Mean difference by *Same sex*	*More than 2 children*	*Number of children*	Mean difference by *Twins-2*	*More than 2 children*	*Number of children*
More than 2 children	0.0600 (0.0016)	—	—	0.0628 (0.0016)	—	—	0.6031 (0.0084)	—	—
Number of children	0.0765 (0.0026)	—	—	0.0836 (0.0025)	—	—	0.8094 (0.0139)	—	—
Worked for pay	−0.0080 (0.0016)	−0.133 (0.026)	−0.104 (0.021)	−0.0053 (0.0015)	−0.084 (0.024)	−0.063 (0.018)	−0.0459 (0.0086)	−0.076 (0.014)	−0.057 (0.011)
Weeks worked	−0.3826 (0.0709)	−6.38 (1.17)	−5.00 (0.92)	−0.3233 (0.0743)	−5.15 (1.17)	−3.87 (0.88)	−1.982 (0.386)	−3.28 (0.63)	−2.45 (0.47)
Hours/week	−0.3110 (0.0602)	−5.18 (1.00)	−4.07 (0.78)	−0.2363 (0.0620)	−3.76 (0.98)	−2.83 (0.73)	−1.979 (0.327)	−3.28 (0.54)	−2.44 (0.40)
Labor income	−132.5 (34.4)	−2208.8 (569.2)	−1732.4 (446.3)	−119.4 (42.4)	−1901.4 (670.3)	−1428.0 (502.6)	−570.8 (186.9)	−946.4 (308.6)	−705.2 (229.8)
ln(*Family income*)	−0.0018 (0.0041)	−0.029 (0.068)	−0.023 (0.054)	−0.0085 (0.0047)	−0.136 (0.074)	−0.102 (0.056)	−0.0341 (0.0223)	−0.057 (0.037)	−0.042 (0.027)

Notes: The samples are the same as in Table 2. Standard errors are reported in parentheses.

Interpretation

- Where does the -0.133 figure come from ?
- How can you interpret it ?
- Where does the -0.104 figure come from ?
- How can you interpret it ?

2SLS results

TABLE 7—OLS AND 2SLS ESTIMATES OF LABOR-SUPPLY MODELS USING 1980 CENSUS DATA

	All women			Married women			Husbands of married women		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Estimation method	OLS	2SLS	2SLS	OLS	2SLS	2SLS	OLS	2SLS	2SLS
Instrument for *More than 2 children*	—	*Same sex*	*Two boys, Two girls*	—	*Same sex*	*Two boys, Two girls*	—	*Same sex*	*Two boys, Two girls*
Dependent variable:									
Worked for pay	−0.176 (0.002)	−0.120 (0.025)	−0.113 (0.025) [0.013]	−0.167 (0.002)	−0.120 (0.028)	−0.113 (0.028) [0.013]	−0.008 (0.001)	0.004 (0.009)	0.001 (0.008) [0.013]
Weeks worked	−8.97 (0.07)	−5.66 (1.11)	−5.37 (1.10) [0.017]	−8.05 (0.09)	−5.40 (1.20)	−5.16 (1.20) [0.071]	−0.82 (0.04)	0.59 (0.60)	0.45 (0.59) [0.030]
Hours/week	−6.66 (0.06)	−4.59 (0.95)	−4.37 (0.94) [0.030]	−6.02 (0.08)	−4.83 (1.02)	−4.61 (1.01) [0.049]	0.25 (0.05)	0.56 (0.70)	0.50 (0.69) [0.71]
Labor income	−3768.2 (35.4)	−1960.5 (541.5)	−1870.4 (538.5) [0.126]	−3165.7 (42.0)	−1344.8 (569.2)	−1321.2 (565.9) [0.703]	−1505.5 (103.5)	−1248.1 (1397.8)	−1382.3 (1388.9) (0.549)
ln(*Family income*)	−0.126 (0.004)	−0.038 (0.064)	−0.045 (0.064) [0.319]	−0.132 (0.004)	−0.051 (0.056)	−0.053 (0.056) [0.743]	—	—	—
ln(*Non-wife income*)	—	—	—	−0.053 (0.005)	0.023 (0.066)	0.016 (0.066) [0.297]	—	—	—

Interpretation

- Is the OLS estimator very different from the IV ? How do you interpret the change ? How is number of children correlated to other determinants of labor market participation than number of children ?
- How do you interpret the result on men ?
- How do the IV standard errors compare to the OLS ?
- Conclusion: should we subsidize the construction of kindergarten ? Will the impact on fertility rate and on women's employment rate be as strong as suggested by a simple OLS regression ?

An efficient tool under some conditions

- IV measure a truly causal effect under the assumption that the instrument is indeed exogenous (as good as random).
- Happens very rarely, except in randomized experiments or with "natural experiments" (sex of the first two children, quarter of birth)...
- Verification of this assumption: common sense if only one instrument, test if more instruments.
- Nevertheless, several issues:
  - weak instruments => huge samples needed to have good estimators
  - the effect measured by IV is not general only local.