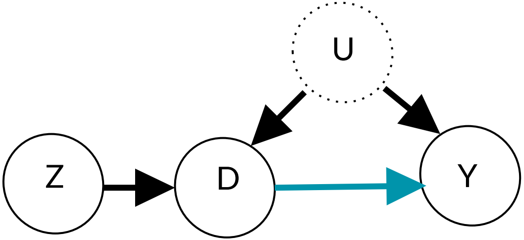
Instrumental variables

Lukáš Lafférs

Matej Bel University, Dept. of Mathematics

Causal graph



- Y is the outcome
- D is a variable of interest (treatment)
- Z is an instrument
- U is an unobserved variable

In this situation, it is not possible to (non-parametrically) identify the causal effect of D on Y.

Things are not completely hopeless though.

Homogenous treatment effects

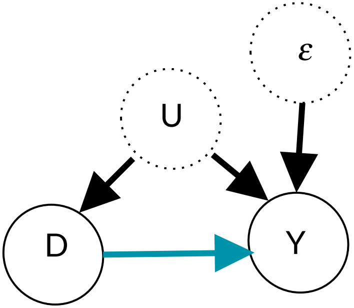
Let us simplify it a little bit.

We will assume:

- homogeneity of the effect
- linearity of the function forms

Thus, we will assume a lot...

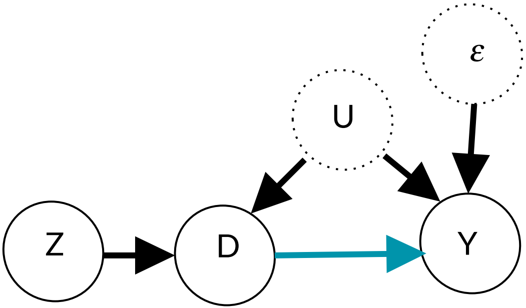
But it makes it possible to proceed in a rather straightforward manner.



- The true relationship is
  $Y_i = \alpha + \delta D_i + \underbrace{\gamma U_i + \varepsilon_i}_{=\eta_i}$
- But U_i is unobserved
  $Y_i = \alpha + \delta D_i + \eta_i$

$$\hat{\delta} = \frac{Cov(Y,D)}{Var(D)} = \frac{E[YD] - E[Y]E[D]}{Var(D)}$$

$$= \frac{E[\alpha D_i + \delta D_i^2 + \gamma U_i D_i + \varepsilon_i D_i] - E[\alpha + \delta D_i + \gamma U_i + \varepsilon_i]E[D]}{Var(D)} = \delta + \gamma \frac{Cov(U,D)}{Var(D)}$$



- The true relationship is
  $Y_i = \alpha + \delta D_i + \underbrace{\gamma U_i + \varepsilon_i}_{=\eta_i}$
- But U_i is unobserved
  $Y_i = \alpha + \delta D_i + \eta_i$
- New variable Z
  Notice: no $Z \rightarrow U$ or $Z \rightarrow Y$

$$
\begin{aligned}
Cov(Y,Z) &= Cov(\alpha + \delta D + \gamma U + \varepsilon, Z) = E[(\alpha + \delta D + \gamma U + \varepsilon)Z] - E[D]E[Z] \\
&= \delta Cov(D,Z) + \gamma \underbrace{Cov(U,Z)}_{=0} + \underbrace{Cov(\varepsilon,Z)}_{=0} \\
&\implies \\
\delta &= \frac{Cov(Y,Z)}{Cov(D,Z)}
\end{aligned}
$$

Exclusion restriction

There are no arrows

- $Z \rightarrow U$
- $Z \rightarrow Y$

This is called an exclusion restriction

Z provides us with the much needed exogenous source of variation

The regression coefficient $\delta = \frac{Cov(Y,Z)}{Cov(D,Z)}$ can be estimated by

$$\hat{\delta} = \frac{\widehat{Cov}(Y,Z)}{\widehat{Cov}(D,Z)} = \frac{\frac{1}{n}\sum_{i=1}^{n}(Z_i - \bar{Z})(Y_i - \bar{Y})}{\frac{1}{n}\sum_{i=1}^{n}(Z_i - \bar{Z})(D_i - \bar{D})}$$

If we assume

$$\begin{aligned} Y_i &= \alpha + \delta D_i + \eta_i \\ D_i &= \beta_0 + \beta_z Z_i + \upsilon_i \end{aligned}$$

Then

$$\hat{\delta} = \frac{\frac{1}{n}\sum_{i=1}^{n}(Z_i - \bar{Z})\overbrace{Y_i}^{=\alpha+\delta D_i+\eta_i}}{\frac{1}{n}\sum_{i=1}^{n}(Z_i - \bar{Z})D_i} = \delta + \overbrace{\frac{\frac{1}{n}\sum_{i=1}^{n}(Z_i - \bar{Z})\eta_i}{\frac{1}{n}\sum_{i=1}^{n}(Z_i - \bar{Z})D_i}}^{\rightarrow_P 0}$$

$$
\begin{aligned}
Y_i &= \alpha + \delta \overbrace{D_i}^{\beta_0+\beta_z Z_i+\upsilon_i} + \eta_i = \underbrace{\alpha_0 + \overbrace{\alpha_Z}^{\delta\cdot\beta_Z} Z_i + \omega_i}_{\text{Reduced form eq.}} \\
D_i &= \underbrace{\beta_0 + \beta_z Z_i + \upsilon_i}_{\text{First stage eq.}}
\end{aligned}
$$

Take a closer look at $\hat{\delta}$

$$
\hat{\delta} = \frac{\widehat{Cov}(Y,Z)}{\widehat{Cov}(D,Z)} = \frac{\frac{\widehat{Cov}(Y,Z)}{\widehat{Var}(Z)}}{\frac{\widehat{Cov}(D,Z)}{\widehat{Var}(Z)}} = \frac{\hat{\alpha}_Z}{\hat{\beta}_Z}
$$

Two-stage least squares

$$\hat{\delta} = \frac{\widehat{Cov}(Y,Z)}{\widehat{Cov}(D,Z)} = \frac{\hat{\beta}_Z \widehat{Cov}(Y,Z)}{\hat{\beta}_Z \widehat{Cov}(D,Z)} = \frac{\widehat{Cov}(Y,\hat{\beta}_Z Z)}{\hat{\beta}_Z^2 \widehat{Var}(Z)}$$

$$= \frac{\widehat{Cov}(Y,\hat{\beta}_Z Z)}{\widehat{Var}(\hat{\beta}_Z Z)} = \cdots = \frac{\widehat{Cov}(Y,\hat{D})}{\widehat{Var}(\hat{D})}$$

where $\hat{D} = \hat{\beta}_0 + \hat{\beta}_Z Z$

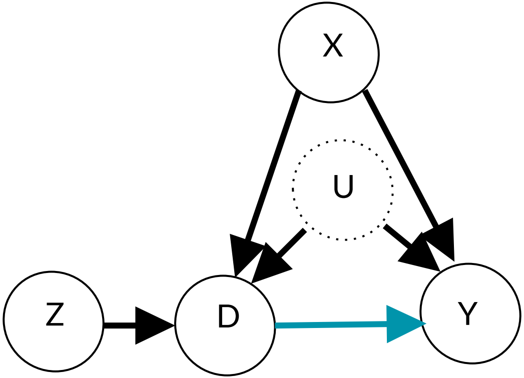
This suggest the following two-stage strategy:

Step 1 Estimate $(\hat{\beta}_0, \hat{\beta}_Z)$ from $D_i = \beta_0 + \beta_z Z_i + \upsilon_i$ and obtain $\hat{D} = \hat{\beta}_0 + \hat{\beta}_Z Z$

Step 2 Plug $\hat{D}$ and estimate $(\hat{\alpha}, \hat{\delta})$ from $Y_i = \alpha + \delta \hat{D}_i + \eta_i$

Such regression coefficient $\hat{\delta}$ will be identical to $\frac{\hat{\alpha}_Z}{\hat{\beta}_Z}$

Additional covariates?



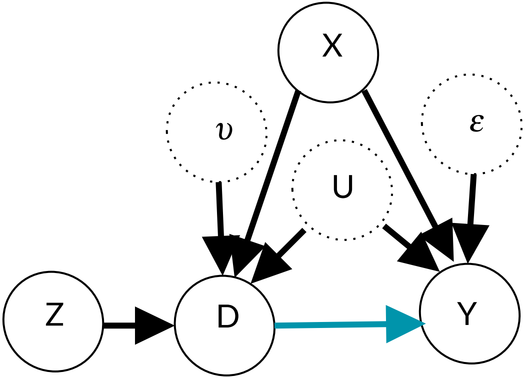
It is important to close all these paths ($D \leftarrow X \rightarrow Y$) too.

Wald estimator

In case of abinary instrument and no covariates, the IV estimator is

$$\hat{\delta}_{IV} = \frac{\hat{E}[Y|Z=1] - \hat{E}[Y|Z=0]}{\hat{E}[D|Z=1] - \hat{E}[D|Z=0]}$$

Additional covariates?



$$\begin{aligned} Y_i &= \alpha + \delta D_i + \delta_X X_i + \overbrace{\delta_U U_i + \varepsilon_i}^{=\eta_i} \\ &= \underbrace{\alpha_0 + \alpha_Z Z_i + \alpha_X X_i + \omega_i}_{\text{Reduced form eq.}} \\ D_i &= \underbrace{\beta_0 + \beta_z Z_i + \beta_X X_i + \upsilon_i}_{\text{First stage eq.}} \end{aligned}$$

Step 1 Estimate $(\hat{\beta}_0, \hat{\beta}_Z, \hat{\beta}_X)$ from $D_i = \beta_0 + \beta_z Z_i + \beta_x X_i + \upsilon_i$ and obtain $\hat{D} = \hat{\beta}_0 + \hat{\beta}_Z Z + \hat{\beta}_X X$

Step 2 Plug $\hat{D}$ and estimate $(\hat{\alpha}, \hat{\delta}, \hat{\delta}_X)$ from $Y_i = \alpha + \delta \hat{D}_i + \delta_X X_i + \eta_i$

Instrument

There are two qualities that the instrument needs to have:

- Validity - instrument Z has no direct effect on Y. It only operates via D. Z needs to be uncorrelated with η_i and therefore with both U_i and ε_i
- Relevance - Z is correlated with D

Where are we now:

- So far, we were unable to non-parametrically identify ATE. We could not close the paths going via confounder U.
- By simplifying a lot, we can at least identify and estimate the regression coefficient δ within a linear model.
- This is a ratio of coefficients from two regression OR we can look at it as two stage estimator
- That is all great as long as the linear model is correct and effects are homogenous.
- Let us see it in action.

Example: children and labor supply

We wish to understand the causal link between the family size and the labor supply.

Do parents of bigger families work more?

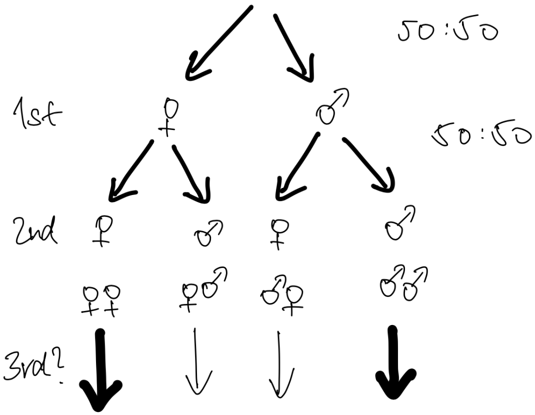
A lot of literature found negative correlation between family size and female labor supply.

How to estimate these? Clearly, the family size is not ”randomly assigned”.

Angrist, Joshua, and William Evans. ”Children and Their Parents’ Labor Supply: Evidence from Exogenous Variation in Family Size.” American Economic Review 88.3 (1998): 450-77.

Example: children and labor supply

Where do we find a proper instrument, that would provide an exogenous variation in the family size?



- Parents have preference for mixed genders
- The gender "assignment" itself is as good as random
- Parents with these kids {(♀, ♀), (♂, ♂)} are more likely to have another one in comparison to parents with {(♀, ♂), (♂, ♀)} kids
- Exogenous variation in the probability of having a third child!

Gender of the first kid does not predict the probability of having the second child.

TABLE 3—FRACTION OF FAMILIES THAT HAD ANOTHER CHILD BY PARITY AND SEX OF CHILDREN

	All women				Married women			
	1980 PUMS (649,887 observations)		1990 PUMS (627,362 observations)		1980 PUMS (410,333 observations)		1990 PUMS (477,798 observations)	
Sex of first child in families with one or more children	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
(1) one girl	0.488	0.694 (0.001)	0.489	0.665 (0.001)	0.485	0.720 (0.001)	0.487	0.698 (0.001)
(2) one boy	0.512	0.694 (0.001)	0.511	0.667 (0.001)	0.515	0.720 (0.001)	0.513	0.699 (0.001)
difference (2) − (1)	—	0.000 (0.001)	—	0.002 (0.001)	—	0.000 (0.001)	—	0.001 (0.001)

Table 3 from Angrist and Evans (1998)

Gender composition predicts the probability of having a third child.

Sex of first two children in families with two or more children	All women: 1980 PUMS (394,835 observations) Fraction of sample	All women: 1980 PUMS Fraction that had another child	All women: 1990 PUMS (380,007 observations) Fraction of sample	All women: 1990 PUMS Fraction that had another child	Married women: 1980 PUMS (254,654 observations) Fraction of sample	Married women: 1980 PUMS Fraction that had another child	Married women: 1990 PUMS (301,588 observations) Fraction of sample	Married women: 1990 PUMS 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)

Table 3 from Angrist and Evans (1998)

Ordinary least squares estimator (for comparison purposes)

$$Y_i = \alpha + \delta D_i + \delta_X X_i + \eta_i$$

Instrumental variable estimation

$$\begin{aligned} Y_i &= \alpha + \delta D_i + \delta_X X_i + \eta_i \\ D_i &= \beta_0 + \beta_z Z_i + \beta_X X_i + \upsilon_i \end{aligned}$$

- Y is one of these
  - worked
  - weeks worked
  - hours/week
  - log family income
  - non-wife income
- D is an indicator of having more than 2 children
- X consists of: age, age at first birth, black indicator, hispanic indicator, boy 1st indicator, boy 2nd indicator
- z is one of these
  - same sex
  - two boys, two girls (as separate instruments)

No covariates - Wald estimates

Table 5—Wald Estimates of Labor-Supply Models

Variable	1980 PUMS: Mean difference by *Same sex*	1980 PUMS: Wald estimate using as covariate: *More than 2 children*	1980 PUMS: Wald estimate using as covariate: *Number of children*	1990 PUMS: Mean difference by *Same sex*	1990 PUMS: Wald estimate using as covariate: *More than 2 children*	1990 PUMS: Wald estimate using as covariate: *Number of children*	1980 PUMS: Mean difference by *Twins-2*	1980 PUMS: Wald estimate using as covariate: *More than 2 children*	1980 PUMS: Wald estimate using as covariate: *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)

Table 5 from Angrist and Evans (1998)

Instrument is relevant

TABLE 6—OLS ESTIMATES OF *MORE THAN 2 CHILDREN* EQUATIONS

Independent variable	All women (1)	(2)	(3)	Married women (4)	(5)	(6)
1980 PUMS						
Boy 1st	—	−0.0080 (0.0015)	0.0001 (0.0021)	—	−0.0111 (0.0018)	−0.0016 (0.0026)
Boy 2nd	—	−0.0081 (0.0015)	—	—	−0.0095 (0.0018)	—
Same sex	0.0600 (0.0016)	0.0617 (0.0015)	—	0.0675 (0.0019)	0.0694 (0.0018)	—
Two boys	—	—	0.0536 (0.0021)	—	—	0.0598 (0.0026)
Two girls	—	—	0.0698 (0.0021)	—	—	0.0789 (0.0026)
With other covariates	no	yes	yes	no	yes	yes
R^2	0.004	0.084	0.084	0.005	0.078	0.078

Table 6 from Angrist and Evans (1998)

With covariates

Magnitude of the effect is smaller than under OLS

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]	—	—	—

Table 7 from Angrist and Evans (1998)

Mechanics

Step 1 Estimate $(\hat{\beta}_0, \hat{\beta}_Z, \hat{\beta}_X)$ from $D_i = \beta_0 + \beta_z Z_i + \beta_x X_i + \upsilon_i$ and obtain $\hat{D} = \hat{\beta}_0 + \hat{\beta}_Z Z + \hat{\beta}_X X$

Step 2 Plug $\hat{D}$ and estimate $(\hat{\alpha}, \hat{\delta}, \hat{\delta}_X)$ from $Y_i = \alpha + \delta \hat{D}_i + \delta_X X_i + \eta_i$

Can be translated as

Step 1 Regress D on all sources of exogenous variation (Z and X)

Step 2 Regress Y on the predicted values $\hat{D}$ of D and exogenous variables X (not instruments!)

Mechanics (it is a simple projection)

$\boldsymbol{X} = [\mathbf{1}, X, D] \quad \boldsymbol{Z} = [\mathbf{1}, X, Z] \quad y_i = \boldsymbol{X}_i \beta + e_i$

$\hat{\beta}_{OLS} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T Y \not\to_P \beta$
because $E(\boldsymbol{X}^T e) \neq 0$

Regress all the columns of $\boldsymbol{X}$ onto $\boldsymbol{Z}$ to obtain $\hat{\boldsymbol{X}}$
$\hat{\boldsymbol{X}} = \boldsymbol{Z}(\boldsymbol{Z}^T \boldsymbol{Z})^{-1} \boldsymbol{Z}^T \boldsymbol{X} = P_{\boldsymbol{Z}} \boldsymbol{X}$
(note that projecting X on $\boldsymbol{Z}$ will give us the same X because it is in $\boldsymbol{Z}$!)

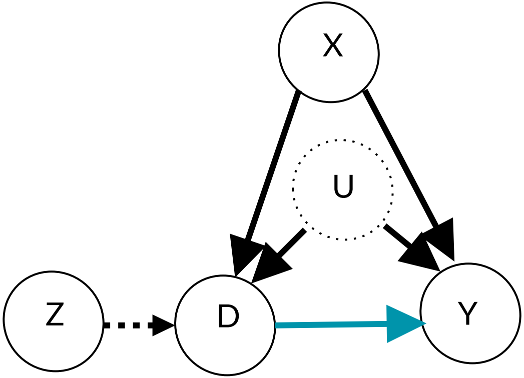
Regress y on $\hat{\boldsymbol{X}}$
$\hat{\beta}_{IV} = (\hat{\boldsymbol{X}}^T \hat{\boldsymbol{X}})^{-1} \hat{\boldsymbol{X}}^T y = ((P_{\boldsymbol{Z}} \boldsymbol{X})^T P_{\boldsymbol{Z}} \boldsymbol{X})^{-1} (P_{\boldsymbol{Z}} \boldsymbol{X})^T y = (\boldsymbol{X}^T \underbrace{P_{\boldsymbol{Z}}}_{= P_{\boldsymbol{Z}}^T P_{\boldsymbol{Z}}} \boldsymbol{X})^{-1} \boldsymbol{X}^T P_{\boldsymbol{Z}} y$

Careful with the standard errors

The second-stage regression does not give you the correct standard errors. (It ignores the first stage uncertainty).

Notice that IV estimator is weighted least squares estimator:
$\hat{\beta}_{IV} = (\hat{\boldsymbol{X}}^T \hat{\boldsymbol{X}})^{-1} \hat{\boldsymbol{X}}^T y = (\boldsymbol{X}^T P_{\boldsymbol{Z}} \boldsymbol{X})^{-1} \boldsymbol{X}^T P_{\boldsymbol{Z}} y$
and thus $\hat{\sigma}^2 (\boldsymbol{X}^T P_{\boldsymbol{Z}} \boldsymbol{X})^{-1}$ is a consistent estimator of covariance matrix of $\hat{\beta}_{IV}$ under homoscedasticity.

Weak instruments



We relied on the fact that there exists this connection: $Z \to D$

But what if the link is only weak?

Weak instruments

So what if the correlation is very small(?)

$$\hat{\delta} = \underbrace{\frac{\widehat{Cov}(Y,Z)}{\widehat{Cov}(D,Z)}}_{\text{very small}} = \frac{\frac{\widehat{Cov}(Y,Z)}{\widehat{Var}(Z)}}{\frac{\widehat{Cov}(D,Z)}{\widehat{Var}(Z)}} = \frac{\hat{\alpha}_Z}{\hat{\beta}_Z}$$

Then the $\hat{\beta}_Z$ is very imprecisely estimated. And this leads to an imprecise estimator for $\hat{\delta}$ itself.

Weak instruments

$$\hat{\delta} = \delta + \frac{\overbrace{\frac{1}{n}\sum_{i=1}^{n}(Z_i - \bar{Z})\eta_i}^{\to_P 0???}}{\underbrace{\frac{1}{n}\sum_{i=1}^{n}(Z_i - \bar{Z})D_i}_{\text{very small}}}$$

Even a tiny small deviation from the exogeneity $Cov(Z, \eta) = 0$ may severely bias our estimator(!)

This is a huge deal.

Bound, John, David A. Jaeger, and Regina M. Baker. "Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak." Journal of the American statistical association 90.430 (1995): 443-450.

Weak instruments

Luckily, we can check if we have this problem simply by looking at the first stage.

Common rule of thumb is to have the value of F-statistic from the first stage regression at least 10.

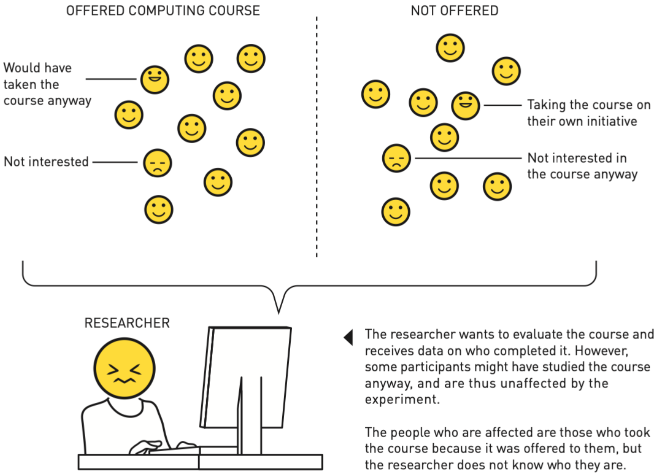
There is a huge strain of literature on weak instruments, many weak instruments etc.

Older Survey: Stock, James H., Jonathan H. Wright, and Motohiro Yogo. "A survey of weak instruments and weak identification in generalized method of moments." Journal of Business & Economic Statistics 20.4 (2002): 518-529.

Newer survey Andrews, Isaiah, James H. Stock, and Liyang Sun. "Weak instruments in instrumental variables regression: Theory and practice." Annual Review of Economics 11 (2019): 727-753.

Statistical Inference: Staiger, Douglas O., and James H. Stock. "Instrumental variables regression with weak instruments." (1994).

Heterogenous effects



https://www.nobelprize.org/uploads/2021/10/fig4_ek_en_21_LATE.pdf

Heterogenous effects

A natural question to ask is the following:

Do all people have the same effect from the treatment?

If not, who are these people who benefit from the treatment?

Interpretation

We now drop the linearity assumption and consider binary treatment and binary instrument.

Every individual i may have her own effect $\delta_i = Y_i(1) - Y_i(0)$ depending on the treatment

Every individual i may also react different in terms of treatment $D_i(1) - D_i(1)$ on the instrument

- Z - randomly offered training
- D - actual training
- Y - outcome

always-taker $D_i(1) = 1$ and $D_i(0) = 1$
complier $D_i(1) = 1$ and $D_i(0) = 0$
defier $D_i(1) = 0$ and $D_i(0) = 1$
never-taker $D_i(1) = 0$ and $D_i(0) = 0$

Denote $Y_i(d, z)$ as a potential outcome under $D_i = d$ and $Z_i = z$.

If

- Instrument is independent of potential outcomes: $(Y_i(D_i(1),1).Y_i(D_i(0),0),D_i(1),D_i(0)) \perp\!\!\!\perp Z_i$
- Exclusion restriction: $Y_i(d) \equiv Y_i(d,1) = Y_i(d,0)$
- Relevance restriction: $E[D_i(1) - D_i(0)] \neq 0$
- Monotonicity: $D_i(1) \geq D_i(0)$
- Stable Unit Treatment Value Assumption: There are no interaction between individuals and there is no hidden variation in the treatment

then

$$\delta_{IV} = \frac{E[Y|Z=1] - E[Y|Z=0]}{E[D|Z=1] - E[D|Z=0]} = \underbrace{E[Y(1) - Y(0)|D(1) > D(0)]}_{\text{Local average treatment effect}}$$

Imbens, G. W. and Angrist, J. D. (1994). Identication and Estimation of Local Average Treatment Effects. Econometrica

Proof

$$E[Y|Z=1] \underbrace{=}_{\text{exclusion}} E[Y(0)+(Y(1)-Y(0))D|Z=1] \underbrace{=}_{\text{Ind.}} E[Y(0)+(Y(1)-Y(0))D(1)]$$

and also

$$E[Y|Z=0]=E[Y(0)+(Y(1)-Y(0))D(0)]$$

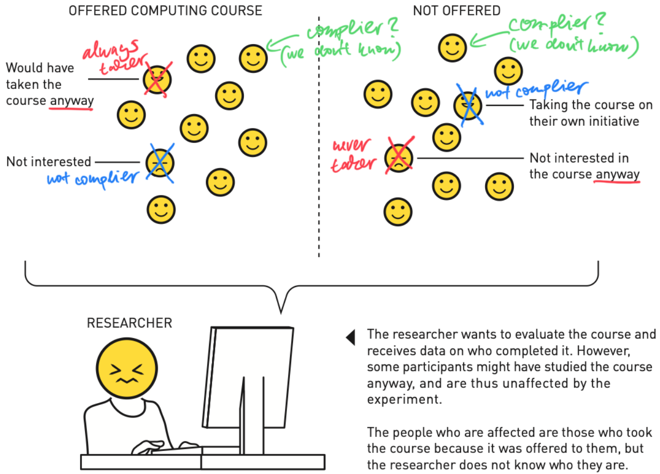
so

$$\begin{aligned} E[Y|Z=1]-E[Y|Z=0] &= E[(Y(1)-Y(0))(D(1)-D(0))] \\ &\underbrace{=}_{\text{mono}} E[(Y(1)-Y(0))|D(1)>D(0)]P(D(1)>D(0)) \end{aligned}$$

Similarly

$$E[D|Z=1]-E[D|Z=0]=E[D(1)-D(0)]=P(D(1)>D(0))$$

Effects on the compliers



The researcher wants to evaluate the course and receives data on who completed it. However, some participants might have studied the course anyway, and are thus unaffected by the experiment.

The people who are affected are those who took the course because it was offered to them, but the researcher does not know who they are.

Effects on the compliers

- LATE interpretation is specific for the instrument
- no restrictions were placed on the homogeneity of the effects
- no linearity was assumed

Extensions:

- Further discussions: Angrist, J. D., Imbens, G. W. and Rubin, D. B. (1996). Identification of Causal Effects Using Instrumental Variables. Journal of the American Statistical Association.
- Multiple valued treatment: Angrist, Joshua D., and Guido W. Imbens. ”Two-stage least squares estimation of average causal effects in models with variable treatment intensity.” Journal of the American statistical Association 90.430 (1995): 431-442.
- Non-parametric LATE with covariates: Frölich, Markus. ”Nonparametric IV estimation of local average treatment effects with covariates.” Journal of Econometrics 139.1 (2007): 35-75.

Further applications

- Returns to schooling - Quarter of birth instrument (Angrist and Krueger, 1991)
- Returns to schooling - Nearby college instrument (Card, 1995)
- Returns to schooling - Different instruments (Ichino and Winter-Ebmer, 1999)
- Classroom size - Legislative rule as instrument (Angrist and Lavy 1999)
- Effect of military service on labor market outcomes - Draft lottery instrument (Angrist, 1990)
- Impact of institutions on economic growth - Mortality instrument (Acemoglu, Johnson and Robinson, 2001), Comment (Albouy, 2012), Reply (AJR, 2012)
- Impact of economic conditions on prob. of a conflict - rainfall instrument (Miguel, Satyanath and Segenti, 2004)

Further applications

- Demand for fish - Weather as an IV (Angrist, Graddy and Imbens)
- Childbearing on labor supply - twin births as a natural experiment (Jacobsen, Pearce and Rosenbloom. 1999) and (Black, Devereux and Salvanes, 2015)
- Using economic theory to estimate supply and demand curves using variation in a single tax rate(!) (Zoutman, Gavrilova and Hopland. 2018)
- Parental Meth Abuse and Foster Care - use supply shock on meth market as instrument (Cunningham and Finlay, 2013)

Measurement error

Suppose that X is measured with error:

$$Y_i = \beta_0 + \beta_X \underbrace{(X_i^* + u_i)}_{X_i} + \varepsilon_i$$

$$\hat{\beta}_X = \frac{\widehat{Cov}(X,Y)}{\widehat{Var}(X)} = \frac{\widehat{Cov}(X^* + u, \beta_0 + \beta_X(X^* + u) + \varepsilon)}{\widehat{Var}(X^* + u)} \to_P \beta_X \frac{\sigma_X^2}{\sigma_X^2 + \sigma_u^2}$$

which is attenuated even if u_i is uncorrelated with both X_i^* and ε_i

AJR 2001

- Institutions - with more secure property rights people will invest more in physical and human capital. Also includes indpendent judiciary, equal access to education and ensuring civil liberties
- Do institutions matter? well, they do: North/South Korea, West/East Germany.
- Different colonization policies: extractive (Kongo) vs strong property rights (Australia, Canada, USA)
- Higher mortality made it more difficult to set up settlements with strong property rights
- Settler mortality $\rightarrow$ Settlements $\rightarrow$ Early institutions $\rightarrow$ Current institutions $\rightarrow$ Current performance

Reduced form

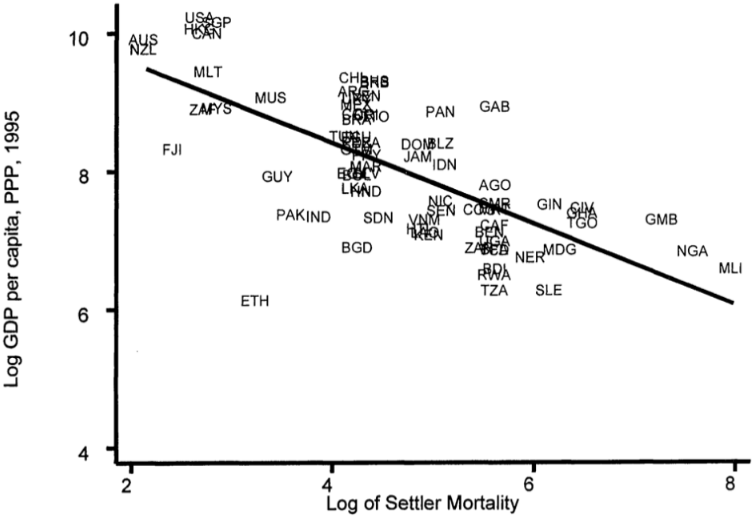


FIGURE 1. REDUCED-FORM RELATIONSHIP BETWEEN INCOME AND SETTLER MORTALITY

AJR 2001

- Exclusion restriction: mortality more that 100yrs ago have no direct impact on GDP per capita today (apart the channel via institutions). Why? Mortality mainly due to malaria and yellow fever.
- Insensitive to outliers (USA, Canada, NZ, Australia)
- Africa dummy and distance to equator insignificant
- Results robust to different covariates added: identify of main colonizer, climate, religion, geography, natural resources, current disease. (in DAG language: closing all the backdoor paths)

TABLE 1—DESCRIPTIVE STATISTICS

	Whole world	Base sample	By quartiles of mortality (1)	(2)	(3)	(4)
Log GDP per capita (PPP) in 1995	8.3	8.05	8.9	8.4	7.73	7.2
	(1.1)	(1.1)				
Log output per worker in 1988 (with level of United States normalized to 1)	−1.70	−1.93	−1.03	−1.46	−2.20	−3.03
	(1.1)	(1.0)				
Average protection against expropriation risk, 1985–1995	7	6.5	7.9	6.5	6	5.9
	(1.8)	(1.5)				
Constraint on executive in 1990	3.6	4	5.3	5.1	3.3	2.3
	(2.3)	(2.3)				
Constraint on executive in 1900	1.9	2.3	3.7	3.4	1.1	1
	(1.8)	(2.1)				
Constraint on executive in first year of independence	3.6	3.3	4.8	2.4	3.1	3.4
	(2.4)	(2.4)				
Democracy in 1900	1.1	1.6	3.9	2.8	0.19	0
	(2.6)	(3.0)				
European settlements in 1900	0.31	0.16	0.32	0.26	0.08	0.005
	(0.4)	(0.3)				
Log European settler mortality	n.a.	4.7	3.0	4.3	4.9	6.3
		(1.1)				
Number of observations	163	64	14	18	17	15

Table 1 in AJR 2001

Model

income per capita

$$\text{(1)} \qquad \log y_i = \mu + \alpha R_i + \mathbf{X}_i'\gamma + \varepsilon_i,$$

protection against expropriation

log M_i not here (exclusion)

$$\text{(5)} \qquad R_i = \zeta + \beta \log M_i + \mathbf{X}_i'\delta + \upsilon_i,$$

settler mortality rate

AJR 2001

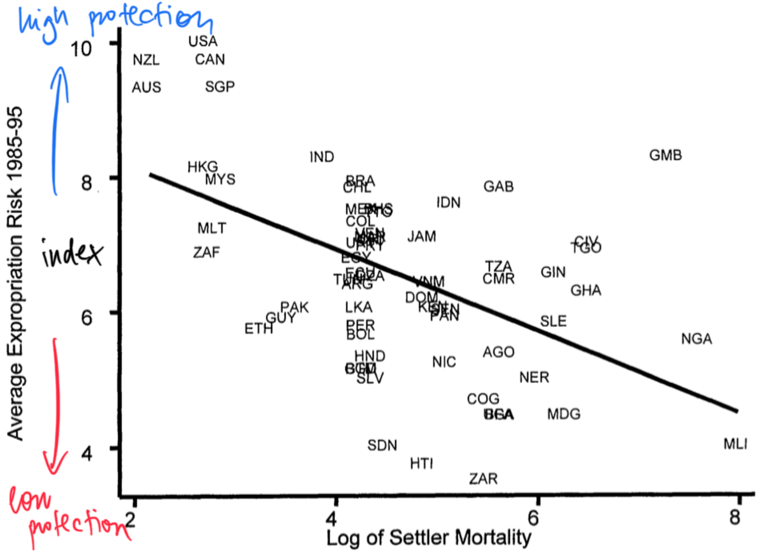


FIGURE 3. FIRST-STAGE RELATIONSHIP BETWEEN SETTLER MORTALITY AND EXPROPRIATION RISK

Fig 3 in AJR 2001

IV estimates

	Base sample (1)	Base sample (2)	Base sample without Neo-Europes (3)	Base sample without Neo-Europes (4)	Base sample without Africa (5)	Base sample without Africa (6)	Base sample with continent dummies (7)	Base sample with continent dummies (8)	Base sample, dependent variable is log output per worker (9)
Panel A: Two-Stage Least Squares									
Average protection against expropriation risk 1985–1995	0.94 (0.16)	1.00 (0.22)	1.28 (0.36)	1.21 (0.35)	0.58 (0.10)	0.58 (0.12)	0.98 (0.30)	1.10 (0.46)	0.98 (0.17)
Latitude		−0.65 (1.34)		0.94 (1.46)		0.04 (0.84)		−1.20 (1.8)	
Asia dummy							−0.92 (0.40)	−1.10 (0.52)	
Africa dummy							−0.46 (0.36)	−0.44 (0.42)	
"Other" continent dummy							−0.94 (0.85)	−0.99 (1.0)	

effects positive

Table 4 in AJR 2001

First stage

Panel B: First Stage for Average Protection Against Expropriation Risk in 1985–1995

Log European settler mortality	−0.61	−0.51	−0.39	−0.39	−1.20	−1.10	−0.43	−0.34	−0.63
	(0.13)	(0.14)	(0.13)	(0.14)	(0.22)	(0.24)	(0.17)	(0.18)	(0.13)
Latitude		2.00		−0.11		0.99		2.00	
		(1.34)		(1.50)		(1.43)		(1.40)	
Asia dummy							0.33	0.47	
							(0.49)	(0.50)	
Africa dummy							−0.27	−0.26	
							(0.41)	(0.41)	
"Other" continent dummy							1.24	1.1	
							(0.84)	(0.84)	
R^2	0.27	0.30	0.13	0.13	0.47	0.47	0.30	0.33	0.28

First stage coeffs

Table 4 in AJR 2001

OLS

OLS coeffs < IV coeffs

Panel C: Ordinary Least Squares

Average protection against expropriation risk 1985–1995	0.52 (0.06)	0.47 (0.06)	0.49 (0.08)	0.47 (0.07)	0.48 (0.07)	0.47 (0.07)	0.42 (0.06)	0.40 (0.06)	0.46 (0.06)
Number of observations	64	64	60	60	37	37	64	64	61

Table 4 in AJR 2001

This is compatible with attenuation bias explanation.

Recap (what have we seen)

- Linear regression
- Maximum likelihood
- Bootstrap
- Causality
- Selection on observables
- IV (...here we are)

Recap IV

- what is IV
- examples and causal graph
- LATE
- AJR
- CF (...here we are)

To do...

- Regression discontinuity design
- Difference in differences
- Synthetic control method
- Machine learning essentials
- Causal machine learning

Example: Meth, Parents and Foster Care (Cunningham and Finley, 2013)

- effect of drug abuse on parenting
- In 1994 - regulation on ephedrine $\rightarrow$ more difficult to produce meth

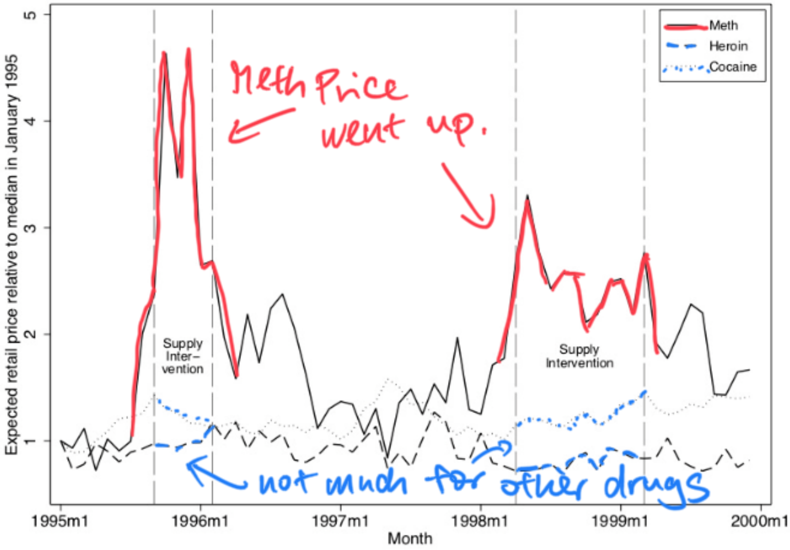


Fig 3 from Cunningham and Finley (2013)

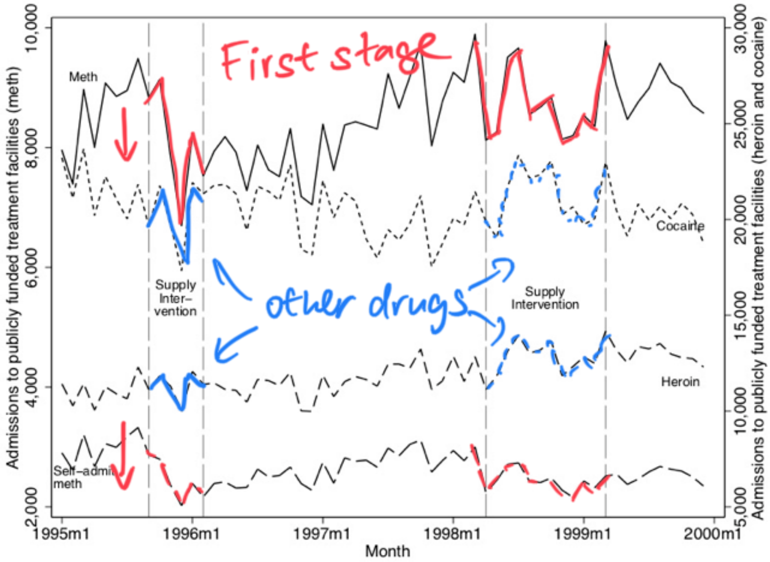


Fig 4 from Cunningham and Finley (2013)

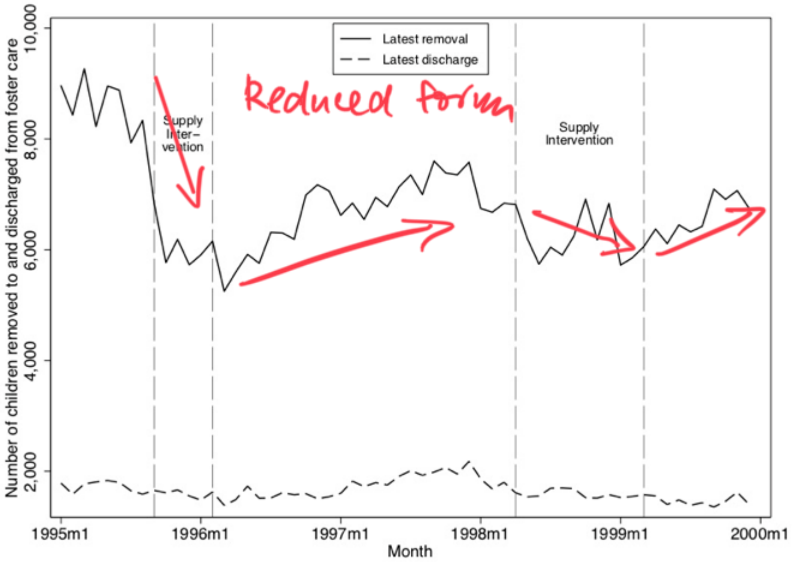


Fig 5 from Cunningham and Finley (2013)

$$\log(\text{self-referred meth treatment})_{st} = \alpha_0 + \alpha_1 \text{price deviation}_t + \alpha_2 \mathbf{X}_{st} + \gamma_s + \phi_t + \tau_{st} + u_{st},$$

D, Z, FIRST STAGE

$$\log(\text{foster care})_{st} = \beta_0 + \beta_1 \times \log(\text{self-referred meth treatment})_{st} + \beta_2 \mathbf{X}_{st} + \delta_s + \lambda_t + \omega_{st} + e_{st},$$

Y, D, SECOND STAGE

- *price deviation*$_t$ the deviation in the expected price of meth from its trend line during precursor regulations and equals zero otherwise
- s - state
- t - specific month
- γ_s, δ_s - state fixed effects
- ϕ_s, λ_t - month-of-year fixed effects
- t_{st}, ω_{st} - state specific linear time trends
- X_{st} - log of state population of whites aged 0-19, 15-49, cigarette tax, state unemployment rate, log of alcohol treatment cases for whites

Covariates	Log Latest Entry into Foster Care OLS (1)	2SLS (2)
Log self-referred meth treatment rate	0.01	1.54***
	(0.02)	(0.59)
Unemployment rate	−0.06**	−0.00
	(0.02)	(0.05)
Cigarette tax per pack	−0.01	0.02
	(0.10)	(0.17)
Log alcohol treatment rate	−0.04	−1.26***
	(0.03)	(0.46)
Log population 0–19 year old	3.68	2.25
	(2.59)	(3.60)
Log population 15–49 year old	−15.48***	−10.61*
	(5.44)	(6.19)
Month-of-year fixed effects	x	x
State fixed effects	x	x
State linear time trends	x	x
First stage		
Price deviation instrument		−0.0005***
		(0.0001)
F-statistic for IV in first stage		17.60
R^2	0.864	
N	1,343	1,343

Part of Table 3 from Cunningham and Finley (2013)

Overidentifying restrictions test

Z may be multidimensional.

Two stage least squares procedure still can be used.

Say we have 2 instruments: Under instrument exogeneity, both of them are fine and hence $\hat{\beta}_{IV1}$ should be similar to $\hat{\beta}_{IV2}$

Under exogeneity, both Z_1 and Z_2 should have zero coefficients in a regression with residuals (using original X and $\hat{\beta}_{IV}$)

F-statistic that jointly tests this multiplied with m is called J-statistic $\sim \chi^2_q$. Where m is the number of instruments, k is the number of endogenous variables and $q = m - k$ is the number of over-identifying restrictions.

See row `Sargan`-row in summary table of `ivreg`.

Wrap up

- IV approach allows to make use of quasi-experimental variation in the treatment that is induced by the instrument.
- IV provides this exogenous variation
- IV needs to be strong enough otherwise estimates are sensitive
- Under monotonicity condition, results informs us only about a specific subpopulation (compliers).

(*) More on IVs

Testable implications on IVs

- Balke and Pearl (1997) for binary Y - based on linear programming
- Huber and Mellace, (2015) - under LATE assumptions
- Kitagawa, (2021) extends Balke and Peal (1997) results to continuous Y
- Zhang. Tian and Bareinboim (2021) - general algorithm for identification of distributions of counterfactual outcomes

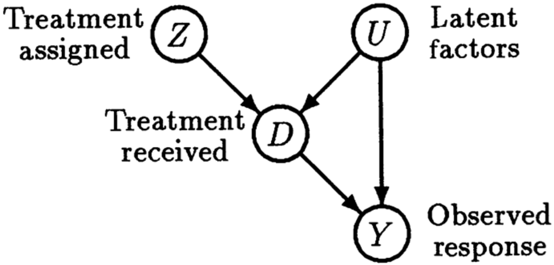


Fig 1 in Balke and Pearl (1997)

If Y, D, Z are discrete, we have that

$$\max_d \sum_y \max_z P(y, d|z) \leq 1$$

Furthermore $ATE = E[Y(1) - Y(0)]$ is bounded.

Thank you for your attention!

References

- Imbens, Guido W., and Joshua D. Angrist. ”Identification and Estimation of Local Average Treatment Effects.” Econometrica 62.2 (1994): 467-475.
- Angrist, Joshua D., and Alan B. Keueger. ”Does compulsory school attendance affect schooling and earnings?.” The Quarterly Journal of Economics 106.4 (1991): 979-1014.
- Angrist, Joshua D. ”Lifetime earnings and the Vietnam era draft lottery: evidence from social security administrative records.” The American Economic Review (1990): 313-336.
- Card, David. ”Using geographic variation in college proximity to estimate the return to schooling.” (1993).
- Acemoglu, Daron, Simon Johnson, and James A. Robinson. ”The colonial origins of comparative development: An empirical investigation.” American economic review 91.5 (2001): 1369-1401.
- Miguel, Edward, Shanker Satyanath, and Ernest Sergenti. ”Economic shocks and civil conflict: An instrumental variables approach.” Journal of political Economy 112.4 (2004): 725-753.
- Ichino, Andrea, and Rudolf Winter-Ebmer. ”Lower and upper bounds of returns to schooling: An exercise in IV estimation with different instruments.” European Economic Review 43.4-6 (1999): 889-901.
- Bound, John, David A. Jaeger, and Regina M. Baker. ”Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak.” Journal of the American statistical association 90.430 (1995): 443-450.
- Stock, James H., Jonathan H. Wright, and Motohiro Yogo. ”A survey of weak instruments and weak identification in generalized method of moments.” Journal of Business & Economic Statistics 20.4 (2002): 518-529.
- Andrews, Isaiah, James H. Stock, and Liyang Sun. ”Weak instruments in instrumental variables regression: Theory and practice.” Annual Review of Economics 11 (2019): 727-753.
- Staiger, Douglas O., and James H. Stock. ”Instrumental variables regression with weak instruments.” (1994).
- Frölich, Markus. ”Nonparametric IV estimation of local average treatment effects with covariates.” Journal of Econometrics 139.1 (2007): 35-75.
- ”Two-stage least squares estimation of average causal effects in models with variable treatment intensity.” Journal of the American statistical Association 90.430 (1995): 431-442.
- Angrist, J. D., Imbens, G. W. and Rubin, D. B. (1996). Identification of Causal Effects Using Instrumental Variables. Journal of the American Statistical Association.

References

- Jacobsen, Joyce P., James Wishart Pearce III, and Joshua L. Rosenbloom. ”The effects of childbearing on married women's labor supply and earnings: using twin births as a natural experiment.” Journal of Human Resources (1999): 449-474.
- Angrist, Joshua D., Kathryn Graddy, and Guido W. Imbens. ”The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish.” The Review of Economic Studies 67.3 (2000): 499-527.
- Zoutman, Floris T., Evelina Gavrilova, and Arnt O. Hopland. ”Estimating both supply and demand elasticities using variation in a single tax rate.” Econometrica 86.2 (2018): 763-771.
- Cunningham, Scott, and Keith Finlay. ”Parental substance use and foster care: Evidence from two methamphetamine supply shocks.” Economic Inquiry 51.1 (2013): 764-782.
- Overidentification test, the very first paper: Sargan, John D. ”The estimation of economic relationships using instrumental variables.” Econometrica: Journal of the Econometric Society (1958): 393-415.
- Balke, Alexander, and Judea Pearl. ”Bounds on treatment effects from studies with imperfect compliance.” Journal of the American Statistical Association 92.439 (1997): 1171-1176.
- Huber, Martin, and Giovanni Mellace. ”Testing instrument validity for LATE identification based on inequality moment constraints.” Review of Economics and Statistics 97.2 (2015): 398-411.
- Kitagawa, Toru. ”The identification region of the potential outcome distributions under instrument independence.” Journal of Econometrics (2021).
- Zhang, Junzhe, Jin Tian, and Elias Bareinboim. ”Partial Counterfactual Identification from Observational and Experimental Data.” arXiv preprint arXiv:2110.05690 (2021).