# Testing the identification of causal effects in observational data

Martin Huber* & Jannis Kueck+

*University of Fribourg, Department of Economics
+University of Hamburg, Faculty of Business Administration

November 7, 2022

# Outline

**In a nutshell**

We want to provide a test for identification of a causal effect based on jointly testing for selection-on-observables and instrument validity in observational data.

### Identifying assumptions in causal inference

- Causal inference typically makes use of identifying assumptions that are deemed to be untestable.
- The selection-on-observables assumption e.g. imposes that the treatment is as good as randomly assigned after controlling for observed covariates.
- Whether the set of covariates is sufficient is typically motivated by economic theory, intuition, or previous empirical findings.
- However, the plausibility of the selection-on-observables assumption often appears 'shaky' and it would be nice to have a statistical test for the identifying assumptions.

## Contribution

- This paper demonstrates that there exists a testable condition for the selection-on-observables assumption in observational data when imposing some causal structure.
- The condition relies on two types of observables, namely covariates to be controlled for and a suspected instrument, and arises if...

Assumption 1 there is no reverse causality from the outcome to the treatment, covariates, or the suspected instrument and from the treatment to the covariates or the suspected instrument,

Assumption 2 the suspected instrument and the treatment are conditionally dependent given the covariates.

# Assumptions on the Causal Structure

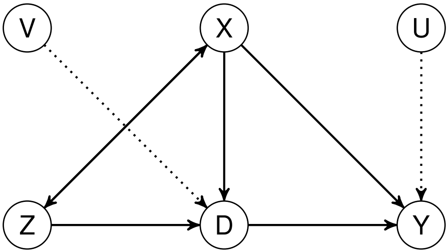
## Notation

- $D$: Treatment.
- $Y$: Outcome.
- $X$: Covariates.
- $Z$: Suspected instrument.
- $V$: Unobservables affecting $D$.
- $U$: Unobservables affecting $Y$.
- $Y(d)$: Potential outcome under treatment $d$.

# Assumptions on the Causal Structure

**Graphical illustration**

Figure: Causal graph satisfying Assumption 1 and Assumption 2



# Identifying Assumptions

**Conditional independence of the treatment**

$$Y(d) \perp\!\!\!\perp D | X \quad \forall d \in \mathscr{D}.$$

- Conditional on $X$, $D$ is not associated with unobserved characteristics affecting the outcome.

**Conditional independence of the instrument**

$$Y(d) \perp\!\!\!\perp Z | X \quad \forall d \in \mathscr{D}.$$

- Conditional on $X$, $Z$ does not directly affect the outcome and is not associated with unobserved characteristics affecting the outcome.

### Theorem

*Conditional on Assumptions 1 and 2, it holds that*

$$Y(d) \perp\!\!\!\perp D | X, \quad Y(d) \perp\!\!\!\perp Z | X \iff Y \perp\!\!\!\perp Z | D = d, X \quad \forall d \in \mathscr{D}.$$

Hence, the testable implication

$$Y \perp\!\!\!\perp Z | D = d, X$$

is necessary and sufficient for the identification of the causal effect when considering potential outcomes $Y(d)$ matching the factual treatment assignment $D = d$.

# Testing Identification

**Testing the identification of average effects**

Theorem

*Conditional on Assumptions 1 and a (stronger) version of Assumption 2 w.r.t. means, it holds that*

$$\begin{aligned} & E[Y(d)|D,X] = E[Y(d)|X], \quad E[Y(d)|Z,X] = E[Y(d)|X] \\ \iff \quad & E[Y|Z,D=d,X] = E[Y|D=d,X] \quad \forall d \in \mathscr{D}. \end{aligned}$$

The testable implication

$$E[Y|Z,D=d,X] = E[Y|D=d,X]$$

is necessary and sufficient for the identification of average effects.

**Testing the identification of average effects**

Denote $\mu(z,d,x) = E[Y|Z=z,D=d,X=x]$. Considering a binary instrument $Z$, conditional mean independence is equivalent to:

$$H_0 : \mu(1,d,x) - \mu(0,d,x) = 0 \quad \forall d \in \mathscr{D}, \text{ and } x \in \mathscr{X}.$$

One approach is to test the condition on average (for a binary $Z$):

$$H_0 : E[\mu(1,D,X) - \mu(0,D,X)] := \Delta = 0.$$

# Testing Identification

**Testing based on DML**

For a binary $Z$, estimate $\Delta$ by double machine learning (DML), see e.g. Chernozhukov et al (2018):

$$\Delta = E\left[\phi(D,X)\right] \tag{1}$$

with

$$\phi(D,X) = \mu(1,D,X) - \mu(0,D,X) + \frac{(Y-\mu(1,D,X))\cdot Z}{p(D,X)} - \frac{(Y-\mu(0,D,X))\cdot(1-Z)}{1-p(D,X)}.$$

- $\mu(Z,D,X)$ and p-score $p(D,X)$ are nuisance terms in Neyman-orthogonal score $\phi(D,X)$.
- Under $n^{-1/4}$-consistency of estimators of $\mu(Z,D,X)$ and $p(D,X)$, the sample analog of (1) is $\sqrt{n}$-consistent.
- Averaging over $X$ may decrease asymptotic power since violations across $X$ can cancel out.

# Testing Identification

**DML-based testing in subsets**

- To increase asymptotic power, test the null in subsamples defined upon values of $X, D$ which are predictive for $\Delta$.
- Procedure:
  1) Split sample randomly into two non-overlapping subsamples.
  2) In the first subsample, estimate $\phi(D,X)$ and predict it as a function of $(D,X)$ using machine learning.
  3) In the second subsample, conduct hypothesis tests using subsets defined as a function of best predictors in $(D,X)$
- Formally, for $L_m$ denoting the $m$th subset, one tests

$$H_0 : E[\mu(1,D,X) - \mu(0,D,X) | (D,X) \in L_m] := \Delta(L_m) = 0.$$

- Using multiplier bootstrap, one can construct joint tests across all subsets based on the norm of $\Delta(L_m)$ for $m = 1, ..., M$.

# Testing Identification

**Testing based on squared differences**

Testing for violations across all values of $D$ and $X$ globally , by verifying the following null hypothesis $H_0$:

$$E[(\mu(1,D,X) - \mu(0,D,X))^2] := \theta = 0.$$

- Testing is based on the Neyman-orthogonal function $(\mu(1,D,X) - \mu(0,D,X))^2 + \zeta$, where $\zeta$ is a mean-zero random variable with bounded variance $\sigma_\zeta^2$ from below.
- $\zeta$ avoids a degenerate variance under $H_0$ (which would violate asymptotic normality).
- Asymptotic normality holds under $H_0$ (but not under a violation of $H_0$).

# Application

**Application to Angrist and Evans (1998) data**

143,410 observations of married white couples with mother's education amounting to 12 years from the 1980 wave of the US Census Public Use Micro Samples.

- $Z$: sex ratio of a mother's first two children,
- $D$: 1 for three or more kids, 0 for two kids,
- $Y$: mother's weeks in employment per year,
- $X$: mother's age, mother's age at first birth, father's age, and father's income.

### Results based on DML and squared differences

Table: Empirical application to the sibling sex instrument

| | est | se | pval |
|---|---|---|---|
| DML | 0.22 | 0.11 | 0.05 |
| Squared difference | 0.5 | 0.02 | 0.00 |

Notes: columns ‘est’, ‘se’, and ‘pval’ provide the estimates, the standard errors and the p-values for $\Delta$ and $\theta$, respectively.

- The tests point to a violation of conditional independence given our limited set of covariates.

## Application

### Results based on DML in subsets

Table: Empirical application to the sibling sex instrument

| splitting variable | subset | est | se | pval | bonf | boot ($l_\infty$) | boot ($l_2$) |
|---|---|---|---|---|---|---|---|
| father's income | in lower half | 0.076 | 0.162 | 0.641 | | | |
| | in upper half | 0.328 | 0.152 | 0.030 | 0.0606 | 0.0588 | 0.0825 |
| treatment D | $D=1$ | 0.080 | 0.175 | 0.646 | | | |
| | $D=0$ | 0.311 | 0.144 | 0.031 | | | |
| mother's age | in lower half | 0.321 | 0.156 | 0.039 | | | |
| | in upper half | 0.139 | 0.159 | 0.382 | | | |
| mother's age (afb) | in lower half | 0.200 | 0.168 | 0.234 | | | |
| | in upper half | 0.226 | 0.148 | 0.128 | | | |
| father's age | in lower half | 0.348 | 0.156 | 0.026 | | | |
| | in upper half | 0.116 | 0.159 | 0.464 | | | |
| father's income | in lower half | 0.076 | 0.162 | 0.641 | | | |
| | in upper half | 0.328 | 0.152 | 0.030 | 0.2555 | 0.2225 | 0.0089 |

Notes: columns 'est', 'se', and 'pval' provide the estimates for $\Delta$ and the respective standard errors and p-values within subsets. The three last columns 'bonf', 'boot ($l_\infty$)', and 'boot ($l_2$)' yield the p-value for the joint significance across all subsets using Bonferroni correction and multiplier bootstrap based on the $l_\infty$-norm and $l_2$-norm, respectively.

- Father's income most predictive variable for heterogeneity in violations of conditional independence.
- Joint tests point to violation of conditional independence.

# Conclusion

- We suggested a testable condition for the joint satisfaction of selection-on-observables and IV validity assumptions, if the instrument and treatment are conditionally dependent.
- We proposed several machine learning-based tests which permit controlling for covariates in a data-driven way when testing.
- We presented an application to the evaluation of the effects of fertility on female labour supply.

# Thank you for your attention!

Our paper is available on arXiv:
`https://arxiv.org/abs/2203.15890`

Contact:
Jannis Kueck
Faculty of Business Administration
University of Hamburg
jannis.kueck@uni-hamburg.de