Testing Instrument Validity in Heterogeneous Treatment Effect Models with Covariates

Anna Krumme[1] Matthias Westphal[2]

[1,2]University of Hagen
[1]TU Dortmund
[2]RWI Essen

EEA Congress 2025 Bordeaux
August 27, 2025

IV validity testing

Instrumental variable (IV) methods are essential for causal inference and can reveal important heterogeneities.

Assumptions for valid IVs are well-known (here for heterogenous treatment effects models):*

- They need to be exogenous
- They only affect the outcome through the treatment
- They affect the treatment only in one direction

Despite availible tests for IV validity, they are not widely conducted in practice. Potential reasons are

- the computational complexity
- difficulties in dealing with many (different types of) covariates

* Note: The relevance condition is directly testable and not necessary for our IV validity test.

Literature & Contribution

There are two test bases for settings with binary instrument and binary treatment:

- mean-based testable implications [Huber and Mellace, 2015] for mean effects (like the LATE)
- density-based conditions [Kitagawa, 2015] for distributional effects (like quantile effects)

Further testing procedures build on the density-based conditions (Mourifié and Wan, 2017; Farbmacher et al., 2022) with extensions for

- multi-valued treatments [Sun, 2023]
- fuzzy RDD [Arai et al., 2022]
- accommodation of a larger number of covariates [Carr and Kitagawa, 2023]

Literature & Contribution

There are two test bases for settings with binary instrument and binary treatment:

- mean-based testable implications [Huber and Mellace, 2015] for mean effects (like the LATE)
- density-based conditions [Kitagawa, 2015] for distributional effects (like quantile effects)

Further testing procedures build on the density-based conditions (Mourifié and Wan, 2017; Farbmacher et al., 2022) with extensions for

- multi-valued treatments [Sun, 2023]
- fuzzy RDD [Arai et al., 2022]
- accommodation of a larger number of covariates [Carr and Kitagawa, 2023]

Contribution

- Addition to Huber and Mellace [2015]: Inclusion of conditioning covariates to the test
- Test procedure with relatively low computational time for linear 2SLS models with many covariates

Setting and Assumptions

In settings with binary treatment D and one binary instrument Z the Wald estimator has a causal interpretation as the local average treatment effect (LATE) under valid assumptions.

$$IV_{Wald} = \frac{E(Y \mid Z = 1) - E(Y \mid Z = 0)}{E(D \mid Z = 1) - E(D \mid Z = 0)}, \tag{2.1}$$

Notation:

- D^z potential treatment choice with a specific value of the instrument $z \in 0, 1$
- Y^{dz} potential outcomes for z and treatment state $d \in 0, 1$

Setting and Assumptions

In settings with binary treatment D and one binary instrument Z the Wald estimator has a causal interpretation as the local average treatment effect (LATE) under valid assumptions.

$$IV_{Wald} = \frac{E(Y \mid Z = 1) - E(Y \mid Z = 0)}{E(D \mid Z = 1) - E(D \mid Z = 0)}, \qquad (2.1)$$

Notation:

- D^z potential treatment choice with a specific value of the instrument $z \in 0, 1$
- Y^{dz} potential outcomes for z and treatment state $d \in 0, 1$

We distinguish 4 (unobserved) types by D^z:

Always-takers(AT):	$D^1 = D^0 = 1$	Never-takers(NT):	$D^1 = D^0 = 0$
Complier(C):	$D^1 = 1 > D^0 = 0$	Defier(DF):	$D^1 = 0 < D^0 = 1$

Setting and Assumptions

Assumtpion 1 (Mean independence):

$E(Y^{dz} \mid Z = 1) = E(Y^{dz}|Z = 0)$ and $E(D^z \mid Z = 1) = E(D^z|Z = 0) \quad \forall d, z\{0, 1\}$

The independence assumption allows a causal interpretation of first stage and the reduced form:

$$\begin{aligned} E(D^1 - D^0) &= \Pr(D^1 = 1) - \Pr(D^0 = 1) = \pi_C - \pi_{DF} \\ E(Y^{d1} - Y^{d0}) &= \pi_{NT}\left[\delta^1_{NT} - \delta^0_{NT}\right] + \pi_{AT}\left[\delta^1_{AT} - \delta^0_{AT}\right] + \pi_C\left[\delta^1_C - \delta^0_C\right] + \pi_{DF}\left[\delta^1_{DF} - \delta^0_{DF}\right] \end{aligned}$$

with π_t being the share and δ^z_t the expected outcome by z of type $t \in \{AT, NT, C, DF\}$

Note: Often the stronger assumption of full independence is assumed: $Y^{d1}, Y^{d0}, D^1, D^0 \perp\!\!\!\perp Z$

Setting and Assumptions

Assumtpion 1 (Mean independence):

$E(Y^{dz} \mid Z = 1) = E(Y^{dz}|Z = 0)$ and $E(D^z \mid Z = 1) = E(D^z|Z = 0) \quad \forall d, z\{0, 1\}$

The independence assumption allows a causal interpretation of first stage and the reduced form:

$$E(D^1 - D^0) = \Pr(D^1 = 1) - \Pr(D^0 = 1) = \pi_C - \pi_{DF}$$

$$E(Y^{d1} - Y^{d0}) = \pi_{NT}\left[\delta^1_{NT} - \delta^0_{NT}\right] + \pi_{AT}\left[\delta^1_{AT} - \delta^0_{AT}\right] + \pi_C\left[\delta^1_C - \delta^0_C\right] + \pi_{DF}\left[\delta^1_{DF} - \delta^0_{DF}\right]$$

with π_t being the share and δ^z_t the expected outcome by z of type $t \in \{AT, NT, C, DF\}$

$\Rightarrow$ Using this notation, we write:

$$\text{IV}_{Wald} = \frac{\pi_{NT}\left[\delta^1_{NT} - \delta^0_{NT}\right] + \pi_{AT}\left[\delta^1_{AT} - \delta^0_{AT}\right] + \pi_C\left[\delta^1_C - \delta^0_C\right] + \pi_{DF}\left[\delta^1_{DF} - \delta^0_{DF}\right]}{\pi_C - \pi_{DF}}$$

Setting and Assumptions

Assumtpion 2 (Mean exclusion restriction):

$E(Y^{d,1}) = E(Y^{d,0})$ for $d \in \{0, 1\}$

The exclusion restriction rules out effects for AT and NT as there are no direct effects of Z on Y:

$$\text{IV}_{Wald} = \frac{\pi_C \left[\delta_C^1 - \delta_C^0\right] + \pi_{DF} \left[\delta_{DF}^1 - \delta_{DF}^0\right]}{\pi_C - \pi_{DF}}$$

Note: Often $Y^{d,1} = Y^{d,0}$ for $d \in \{0, 1\}$ is assumed.

Setting and Assumptions

Assumtpion 2 (Mean exclusion restriction):

$E(Y^{d,1}) = E(Y^{d,0})$ for $d \in \{0, 1\}$

The exclusion restriction rules out effects for AT and NT as there are no direct effects of Z on Y:

$$\text{IV}_{Wald} = \frac{\pi_C \left[\delta_C^1 - \delta_C^0\right] + \pi_{DF}\left[\delta_{DF}^1 - \delta_{DF}^0\right]}{\pi_C - \pi_{DF}}$$

Note: Often $Y^{d,1} = Y^{d,0}$ for $d \in \{0, 1\}$ is assumed.

Assumtpion 3 (Monotonicity):

$Pr(D^1 \geq D^0) = 1$

Monotonicity rules out the existance of defiers, $\pi_{DF} = 0$:

$$\text{IV}_{Wald} = \frac{\pi_C \left[\delta_C^1 - \delta_C^0\right]}{\pi_C} = \delta_C^1 - \delta_C^0$$

Testable Implications

From assumptions 1-3 follows: $\pi_{DF} = 0$, $\delta^1_{AT} = \delta^0_{AT}$ and $\delta^1_{NT} = \delta^0_{NT}$ must hold!

Problem: We can not distinguish the types and their potential outcome means directly in the data!

Testable Implications

From assumptions 1-3 follows: $\pi_{DF} = 0$, $\delta^1_{AT} = \delta^0_{AT}$ and $\delta^1_{NT} = \delta^0_{NT}$ must hold!

Problem: We can not distinguish the types and their potential outcome means directly in the data!

What we know by stratifying the sample by D and Z:

		D = 0	D = 1
Z	0	NT^0, C^0	AT^0, DF^0
	1	NT^1, DF^1	AT^1, C^1

Testable Implications

From assumptions 1-3 follows: $\pi_{DF} = 0$, $\delta^1_{AT} = \delta^0_{AT}$ and $\delta^1_{NT} = \delta^0_{NT}$ must hold!

Problem: We can not distinguish the types and their potential outcome means directly in the data!

What we know by stratifying the sample by D and Z:

		D = 0	D = 1
Z	0	NT^0, C^0	AT^0
	1	NT^1	AT^1, C^1

Without defiers we know $\delta^0_{AT} = E(Y|D=1, Z=0)$ and $\delta^1_{NT} = E(Y|D=0, Z=1)$.

But we need information on δ^1_{AT} and δ^0_{NT}

Testable Implications

We know relation between groups (law of total/iterated expectations):

$$\delta^1_{AT,C} = \frac{\pi_C}{\pi_C + \pi_{AT}}\delta^1_C + \frac{\pi_{AT}}{\pi_C + \pi_{AT}}\delta^1_{AT}$$

$$\delta^0_{NT,C} = \frac{\pi_C}{\pi_C + \pi_{NT}}\delta^0_C + \frac{\pi_{NT}}{\pi_C + \pi_{NT}}\delta^0_{NT}$$

Seminal insight by Imbens and Rubin [1997].

Testable Implications

We know relation between groups (law of total/iterated expectations):

$$\delta^1_{AT,C} = \frac{\pi_C}{\pi_C + \pi_{AT}} \delta^1_C + \frac{\pi_{AT}}{\pi_C + \pi_{AT}} \delta^1_{AT}$$

$$\delta^0_{NT,C} = \frac{\pi_C}{\pi_C + \pi_{NT}} \delta^0_C + \frac{\pi_{NT}}{\pi_C + \pi_{NT}} \delta^0_{NT}$$

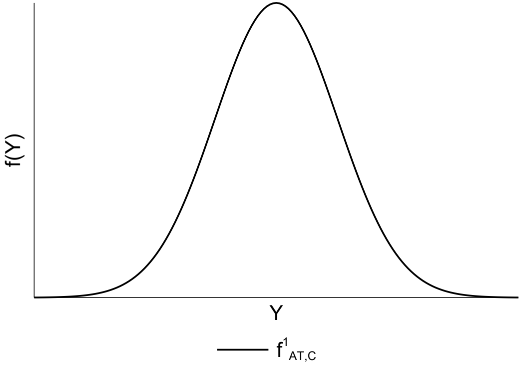
$\Rightarrow$ Relation must hold in expectation and distribution

Seminal insight by Imbens and Rubin [1997].

Testable implications

Identify bounds of unobserved mean potential outcomes (exemplary for $Z = 1$)

Look at the mixed density function ($f^1_{AT,C}(Y)$) and consider extreme case scenarios of the AT^1 group:

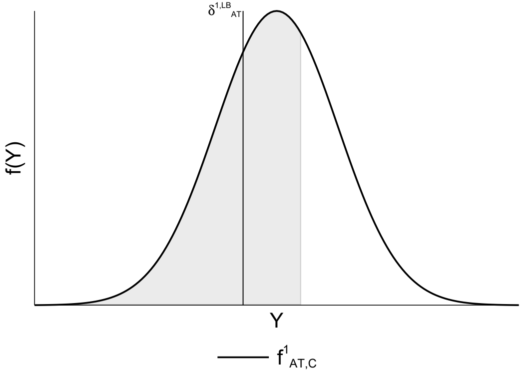


Testable implications

Identify bounds of unobserved mean potential outcomes (exemplary for $Z = 1$)

Look at the mixed density function ($f^1_{AT,C}(Y)$) and consider extreme case scenarios of the AT^1 group:

- share of AT is placed in the lowest ranks of the distribution



Lower bound:

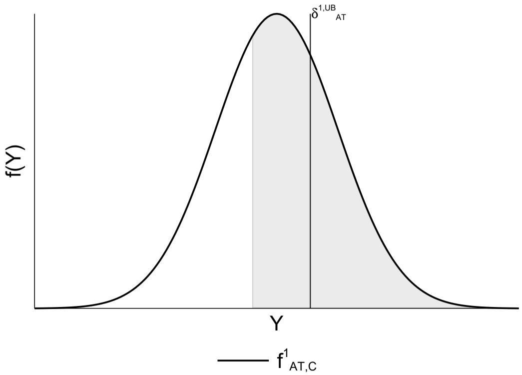
$$\delta^{1,LB}_{AT} = \int_0^{\frac{\pi_{AT}}{\pi_{AT}+\pi_C}} y \, dF(Y = y \mid D = 1, Z = 1)$$

Testable implications

Identify bounds of unobserved mean potential outcomes (exemplary for $Z = 1$)

Look at the mixed density function ($f^1_{AT,C}(Y)$) and consider extreme case scenarios of the AT^1 group:

- share of AT is placed in the highest ranks of the distribution



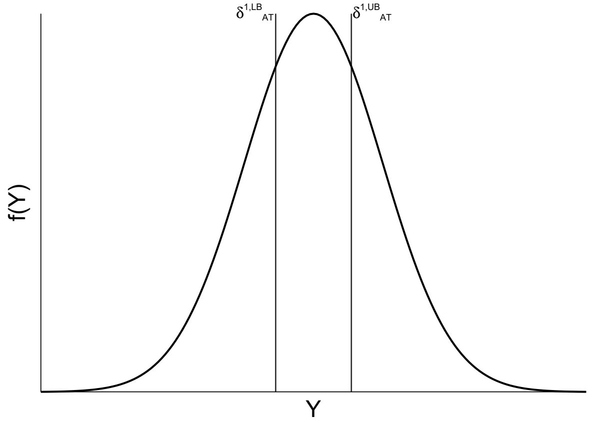
Upper bound:

$$\delta^{1,UB}_{AT} = \int^{1}_{\frac{\pi_{AT}}{\pi_{AT}+\pi_C}} y \, dF(Y = y \mid D = 1, Z = 1)$$

Testable implications

Compare observed potential outcome means to the bounds (exemplary for $Z = 1$)

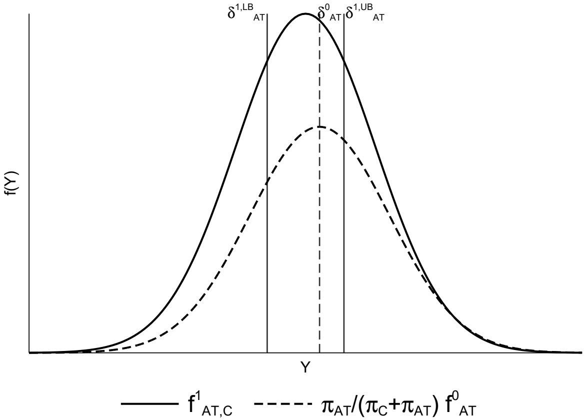
If $\delta^1_{AT} = \delta^0_{AT}$ holds, δ^0_{AT} must lie within the bounds of δ^1_{AT}:



Testable implications

Compare observed potential outcome means to the bounds (exemplary for $Z = 1$)

If $\delta^1_{AT} = \delta^0_{AT}$ holds, δ^0_{AT} must lie within the bounds of δ^1_{AT}:



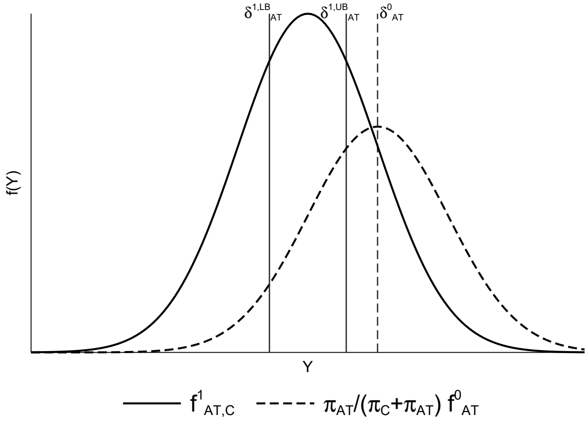
$\delta^{1,LB}_{AT} \leq \delta^0_{AT} \leq \delta^{1,UB}_{AT}$

$\Rightarrow$ No violation detected

Testable implications

Compare observed potential outcome means to the bounds (exemplary for $Z = 1$)

If $\delta^1_{AT} = \delta^0_{AT}$ holds, δ^0_{AT} must lie within the bounds of δ^1_{AT}:



$\delta^{1,LB}_{AT} \leq \delta^0_{AT} \leq \delta^{1,UB}_{AT}$

$\Rightarrow$ Violation detected

Testing

Testable implications (equivalent to Huber and Mellace, 2015):

- if either $\delta^0_{AT} \notin [\delta^{1,LB}_{AT}, \delta^{1,UB}_{AT}]$ or $\delta^1_{NT} \notin [\delta^{0,LB}_{NT}, \delta^{0,UB}_{NT}]$ we can reject IV validity as one of assumptions 1–3 must be violated.

Hypothesis under IV validity:

$$H_0 : \begin{pmatrix} \theta_1 \\ \theta_0 \end{pmatrix} \leq \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

with

$$\theta_1 = \begin{cases} \delta^0_{AT} - \delta^{1,UB}_{AT} & \text{if } \delta^{1,LB}_{AT} < \delta^0_{AT} \\ \delta^{1,LB}_{AT} - \delta^0_{AT} & \text{else.} \end{cases}$$

and

$$\theta_0 = \begin{cases} \delta^1_{NT} - \delta^{0,UB}_{NT} & \text{if } \delta^{0,LB}_{NT} < \delta^1_{NT} \\ \delta^{0,LB}_{NT} - \delta^1_{NT} & \text{else.} \end{cases}$$

Estimation

1. First Stage to get shares: $D_i = \pi_{AT} + \pi_C Z_i + \widetilde{X}_i' \delta + U_{Di}$
   where $\widetilde{X}$ indicates the demeaned covariate vector X

Estimation

1. First Stage to get shares: $D_i = \pi_{AT} + \pi_C Z_i + \widetilde{X}_i'\delta + U_{Di}$
   where $\widetilde{X}$ indicates the demeaned covariate vector X
2. Distribution Regressions to get conditional CDFs:
   - choose evaluation points y within the support of Y

$$\mathbb{1}[Y \leq y] = F^0_{NT,C}(y)\mathbb{1}[D = 0]\mathbb{1}[Z = 0] + F^0_{AT}(y)\mathbb{1}[D = 1]\mathbb{1}[Z = 0] \\ + F^1_{NT}(y)\mathbb{1}[D = 1]\mathbb{1}[Z = 0] + F^1_{AT,C}(y)\mathbb{1}[D = 1]\mathbb{1}[Z = 1] + \widetilde{X}'\lambda + v$$

   $F^0_{NT,C}$, F^0_{AT}, F^1_{NT}, and $F^1_{AT,C}$ measure the share of observations conditional on $D = d$ and $Z = z$ below the threshold y

Estimation

1. First Stage to get shares: $D_i = \pi_{AT} + \pi_C Z_i + \widetilde{X}_i'\delta + U_{Di}$
   where $\widetilde{X}$ indicates the demeaned covariate vector X
2. Distribution Regressions to get conditional CDFs:
   - choose evaluation points y within the support of Y

$$\mathbb{1}[Y \leq y] = F^0_{NT,C}(y)\mathbb{1}[D=0]\mathbb{1}[Z=0] + F^0_{AT}(y)\mathbb{1}[D=1]\mathbb{1}[Z=0] + F^1_{NT}(y)\mathbb{1}[D=1]\mathbb{1}[Z=0] + F^1_{AT,C}(y)\mathbb{1}[D=1]\mathbb{1}[Z=1] + \widetilde{X}'\lambda + v$$

   $F^0_{NT,C}$, F^0_{AT}, F^1_{NT}, and $F^1_{AT,C}$ measure the share of observations conditional on $D = d$ and $Z = z$ below the threshold y
3. Kernel-weighted local polynomial regressions to derive group-specific PDFs from CDFs

Estimation

1. First Stage to get shares: $D_i = \pi_{AT} + \pi_C Z_i + \widetilde{X}_i' \delta + U_{Di}$
   where $\widetilde{X}$ indicates the demeaned covariate vector X
2. Distribution Regressions to get conditional CDFs:
   - choose evaluation points y within the support of Y

$$\mathbb{1}[Y \leq y] = F^0_{NT,C}(y)\mathbb{1}[D=0]\mathbb{1}[Z=0] + F^0_{AT}(y)\mathbb{1}[D=1]\mathbb{1}[Z=0] + F^1_{NT}(y)\mathbb{1}[D=1]\mathbb{1}[Z=0] + F^1_{AT,C}(y)\mathbb{1}[D=1]\mathbb{1}[Z=1] + \widetilde{X}'\lambda + v$$

   $F^0_{NT,C}$, F^0_{AT}, F^1_{NT}, and $F^1_{AT,C}$ measure the share of observations conditional on $D = d$ and $Z = z$ below the threshold y
3. Kernel-weighted local polynomial regressions to derive group-specific PDFs from CDFs
4. Calculate conditional means and bounds ($\delta^0_{AT}, \delta^1_{NT}, \delta^{1,LB}_{AT}, \delta^{1,UB}_{AT}, \delta^{0,LB}_{NT}\delta^{0,UB}_{NT}$) and plug in to get testing parameters (θ_1, θ_0)

Estimation

1. First Stage to get shares: $D_i = \pi_{AT} + \pi_C Z_i + \widetilde{X}_i' \delta + U_{Di}$
   where $\widetilde{X}$ indicates the demeaned covariate vector X
2. Distribution Regressions to get conditional CDFs:
   - choose evaluation points y within the support of Y

$$\mathbb{1}[Y \leq y] = F^0_{NT,C}(y)\mathbb{1}[D=0]\mathbb{1}[Z=0] + F^0_{AT}(y)\mathbb{1}[D=1]\mathbb{1}[Z=0] + F^1_{NT}(y)\mathbb{1}[D=1]\mathbb{1}[Z=0] + F^1_{AT,C}(y)\mathbb{1}[D=1]\mathbb{1}[Z=1] + \widetilde{X}'\lambda + v$$

   $F^0_{NT,C}$, F^0_{AT}, F^1_{NT}, and $F^1_{AT,C}$ measure the share of observations conditional on $D = d$ and $Z = z$ below the threshold y
3. Kernel-weighted local polynomial regressions to derive group-specific PDFs from CDFs
4. Calculate conditional means and bounds $(\delta^0_{AT}, \delta^1_{NT}, \delta^{1,LB}_{AT}, \delta^{1,UB}_{AT}, \delta^{0,LB}_{NT} \delta^{0,UB}_{NT})$ and plug in to get testing parameters (θ_1, θ_0)
5. Derive inference by bootstrapping the θ's and correcting for multiple testing

▸ bootstrap

Simulation

$$Y = X'\beta_X + \beta_D D + \beta_Z Z + U \qquad \text{with } \beta_{X,1} = \beta_{X,2} = \beta_{X,3} = 1 \text{ and } \beta_D = 1$$

$$D = \mathbb{1}[\pi_0 + \pi_1 Z + U_D \geq 0] \qquad \text{with } \pi_0 = \Phi^{-1}(0.45) \text{ and } \pi_1 = \Phi^{-1}(0.55) - \Phi^{-1}(0.45)$$

$$Z = \mathbb{1}[X'\gamma + U_Z \geq 0]$$

$$X = (X_1, X_2, X_3);\ X_j \sim N(0, I)\ \forall j \in \{1, 2, 3\}$$

$$U_Z \sim N(0, 1) \qquad U, U_D \sim N(0, \Sigma) \qquad \text{with } \Sigma = \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix}$$

Cases:

		Independence	
		$\gamma_1 = \gamma_2 = \gamma_3 = 0$	$\gamma_1 = \gamma_2 = \gamma_3 = 0.22$
Exclusion restriction	$\beta_Z = 0$	Unconditionally valid	Valid only with covariates
	$\beta_Z = 1$	Not valid	Not valid

Simulation

	Z and *X* are independent ($\gamma_1 = \gamma_2 = \gamma_3 = 0$)			*Z* depends on *X* ($\gamma_1 = \gamma_2 = \gamma_3 = 0.22$)		
Nominal size:	0.1	0.05	0.01	0.1	0.05	0.01
Exclusion restriction holds: $\beta_Z = 0$						
w/ covariates	0.00	0.00	0.00	0.00	0.00	0.00
w/o covariates	0.00	0.00	0.00	0.73	0.61	0.41
Huber & Mellace (2015)	0.00	0.00	0.00	0.56	0.46	0.24
Exclusion restriction violated: $\beta_Z = 1$						
w/ covariates	0.70	0.64	0.45	0.80	0.75	0.49
w/o covariates	0.64	0.52	0.28	1.00	1.00	1.00
Huber & Mellace (2015)	0.40	0.33	0.13	1.00	1.00	1.00

Notes: N=1000. Rejection rates are based on $S = 100$ simulations. p-values base on $B = 499$ bootstrap repetitions.

Conclusion

We propose an easily implementable testing procedure that allows testing the LATE assumptions conditional on covariates without drastically increasing computation times.

Performing Monte Carlo exercises, we showed that the testing procedure performs well in finite sample sizes.

We applied the test to the draft eligibility and college proximity instruments from the literature.

Current work: Increase number of simulations

References I

Arai, Y., Hsu, Y.-C., Kitagawa, T., Mourifié, I., and Wan, Y. (2022). Testing identifying assumptions in fuzzy regression discontinuity designs. Quantitative Economics, 13(1):1–28.

Carr, T. and Kitagawa, T. (2023). Testing instrument validity with covariates. Papers, arXiv.org.

Farbmacher, H., Guber, R., and Klaassen, S. (2022). Instrument validity tests with causal forests. Journal of Business & Economic Statistics, 40(2):605–614.

Hall, P. and Wilson, S. R. (1991). Two guidelines for bootstrap hypothesis testing. Biometrics, 47(2):757–762.

Huber, M. and Mellace, G. (2015). Testing instrument validity for late identification based on inequality moment constraints. The Review of Economics and Statistics, 97(2):398–411.

Imbens, G. W. and Rubin, D. B. (1997). Estimating outcome distributions for compliers in instrumental variables models. The Review of Economic Studies, 64(4):555–574. Publisher: Wiley-Blackwell.

Kitagawa, T. (2015). A test for instrument validity. Econometrica, 83(5):2043–2063.

MacKinnon, J. G. (2009). Bootstrap Hypothesis Testing, chapter 6, pages 183–213. John Wiley & Sons, Ltd.

Mourifié, I. and Wan, Y. (2017). Testing local average treatment effect assumptions. The Review of Economics and Statistics, 99(2):305–313.

Sun, Z. (2023). Instrument validity for heterogeneous causal effects. Journal of Econometrics, 237(2, Part A):105523.

Šidák, Z. (1967). Rectangular confidence regions for the means of multivariate normal distributions. Journal of the American Statistical Association, 62(318):626–633.

Bootstrap

1. Draw B bootstrap samples of size N randomly from the original sample with replacement: $b \in \{1, 2, ..., B\}$
2. Estimate $\widehat{\theta}_{1,b}$ and $\widehat{\theta}_{0,b}$ within every sample
3. Recenter the parameter from each bootstrap sample, such that $\widetilde{\theta}_{1,b} = \widehat{\theta}_{1,b} - \widehat{\theta}_1$ and $\widetilde{\theta}_{0,b} = \widehat{\theta}_{0,b} - \widehat{\theta}_0$ to increase testing power if bootstrap samples are drawn from populations that do not satisfy H_0 (suggested by Hall and Wilson [1991])
4. To test the constraints of the H_0 against an upper-tailed alternative hypothesis separately, the bootstrap p-values are then given by

$$\begin{aligned} p_{\widehat{\theta}_1} &= \frac{1}{B} \sum_{b=1}^{B} \mathbb{1}[\widetilde{\theta}_{1,b} > \widehat{\theta}_1] \\ p_{\widehat{\theta}_0} &= \frac{1}{B} \sum_{b=1}^{B} \mathbb{1}[\widetilde{\theta}_{0,b} > \widehat{\theta}_0] \end{aligned} \tag{5.1}$$

5. For a joint test we apply the Šidák correction with the joint p-value $\widehat{p} = 1 - (1 - \min(p_{\widehat{\theta}_1}, p_{\widehat{\theta}_0}))^m$ with m being the number of tests [Šidák, 1967]

back

Simulation

	Z and X are independent ($\gamma_1 = \gamma_2 = \gamma_3 = 0$)			Z depends on X ($\gamma_1 = \gamma_2 = \gamma_3 = 0.22$)		
Nominal size:	0.1	0.05	0.01	0.1	0.05	0.01
Exclusion restriction holds: $\beta_Z = 0$						
w/ covariates						
N=250	0.050	0.020	0.010	0.050	0.020	0.005
N=1000	0.000	0.000	0.000	0.000	0.000	0.000
w/o covariates						
N=250	0.085	0.045	0.005	0.495	0.415	0.200
N=1000	0.000	0.000	0.000	0.730	0.610	0.410
Huber & Mellace (2015)						
N=250	0.010	0.000	0.000	0.340	0.125	0.035
N=1000	0.000	0.000	0.000	0.560	0.460	0.240

Notes: Rejection rates are based on S = 100 simulations. p-values base on B = 499 bootstrap repetitions.

Simulation

	Z and X are independent ($\gamma_1 = \gamma_2 = \gamma_3 = 0$)			Z depends on X ($\gamma_1 = \gamma_2 = \gamma_3 = 0.22$)		
Nominal size:	0.1	0.05	0.01	0.1	0.05	0.01
Exclusion restriction violated: $\beta_Z = 1$						
w/ covariates						
N=250	0.565	0.485	0.340	0.660	0.575	0.41
N=1000	0.700	0.640	0.450	0.800	0.750	0.490
w/o covariates						
N=250	0.455	0.370	0.250	0.975	0.955	0.865
N=1000	0.640	0.520	0.280	1.000	1.000	1.000
Huber & Mellace (2015)						
N=250	0.235	0.185	0.055	0.905	0.845	0.555
N=1000	0.460	0.330	0.130	1.000	1.000	1.000

Notes: Rejection rates are based on $S = 100$ simulations. p-values base on $B = 499$ bootstrap repetitions.

Empirical applications

Application 1: Evaluation of military service on civilian earnings using the draft eligibility instrument

- Instrument: Vietnam era draft lottery
- Treatment: Veteran status
- Independence & Monotonicity likely to hold
- Drafted individuals might change years of education or migration choice to avoid serving in the army $\Rightarrow$ Exclusion restriction violated

Empirical applications - Application 1

Data:

- 1984 Survey of Income and Program Participation (SIPP) including birth cohorts 1944-1953

Empirical Setting:

$$D_i \quad = \pi_0 + \pi_1 Z_i + X_1' \gamma + v_i \tag{5.2}$$

$$Y_i \quad \alpha + \beta \hat{D}_i + X_i' \delta + \epsilon_i \tag{5.3}$$

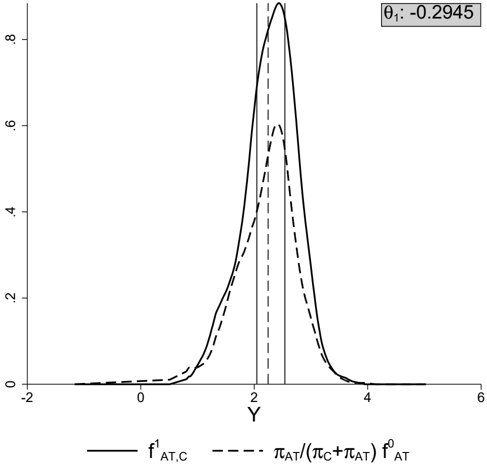
D_i Veteran status

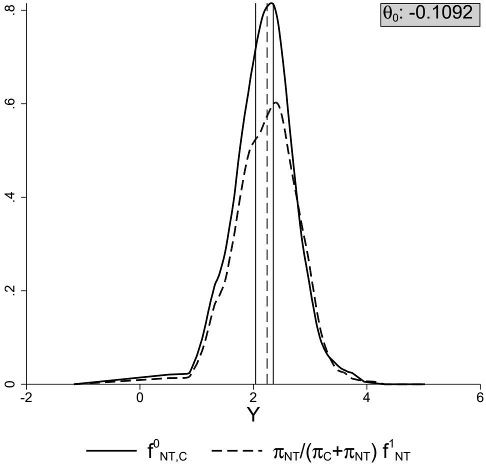
Z_i Draft eligible

Y_i Log of weekly wage

X_i Birth-cohort dummies and race

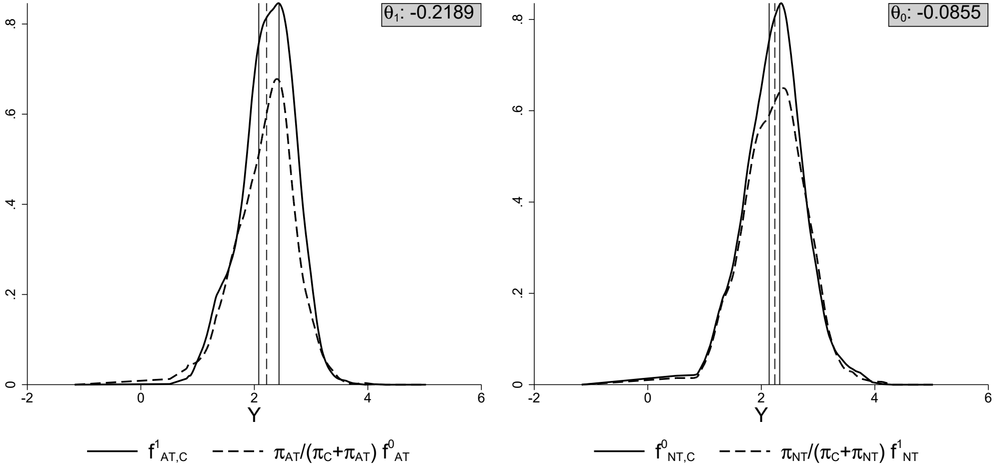
without covariates:





Empirical applications - Application 1

with covariates:



Empirical applications

Application 2: Monetary returns to college education using the college proximity instrument

- Instrument: Four-year college in local labor market
- Treatment: Four-year college degree
- Monotonicity likely to hold, but random assignment unlikely without covariates
- Colleges might increase wages in the local labor market also for NT1, while NT0 stay in their labor markets with lower wage levels $\Rightarrow$ Exclusion restriction potentially violated

Empirical applications - Application 2

Data:

- National Longitudinal Survey of Young Men (NLSYM) including the years 1966-1981 (men aged 14-24 in 1966)

Empirical Setting:

$$D_i = \pi_0 + \pi_1 Z_i + X_1' \gamma + v_i \quad (5.4)$$

$$Y_i \quad \alpha + \beta \hat{D}_i + X_i' \delta + \epsilon_i \quad (5.5)$$

D_i Years of education $\geq$ 16 (1976)

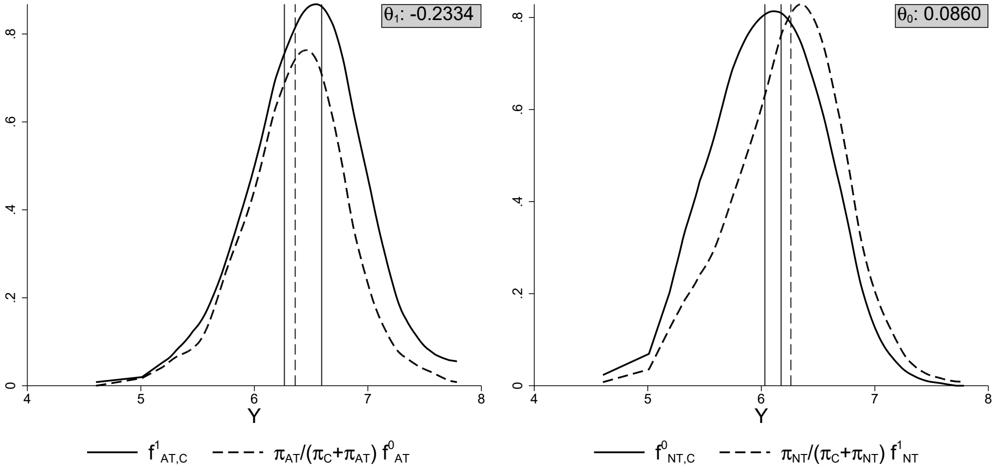
Z_i Four-year college in local labor market (1966)

Y_i Log of weekly wage (1976)

X_i Race, region (1966), metropolitan area (1966&1976), living in south (1976), parents education, family background, experience

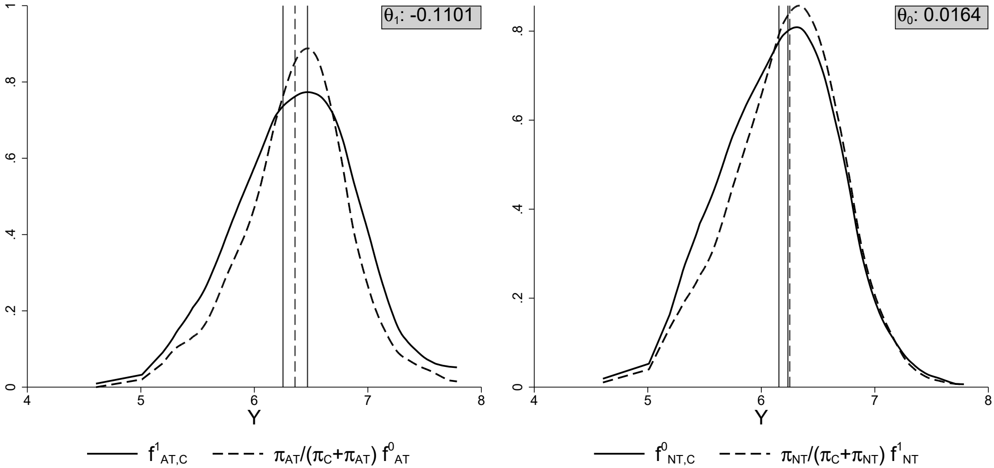
Empirical applications - Application 2

without covariates:



Empirical applications - Application 2

with covariates:



Empirical applications - Application 1 & 2

	Draft lottery		College proximity	
	w/o covariates	w/ covariates[a]	w/o covariates	w/ covariates[b]
θ_1	-0.295	-0.219	-0.233	-0.110
$p_{\hat{\theta}_1}$	1.000	1.000	1.000	0.996
θ_0	-0.109	-0.086	0.086	0.016
$p_{\hat{\theta}_0}$	1.000	1.000	0.002	0.323
Šidák corrected $\hat{p}$	1.000	1.000	0.004	0.541
Shares				
π_C	0.139	0.088	0.069	0.035
π_{AT}	0.265	0.288	0.225	0.248
π_{NT}	0.596	0.623	0.707	0.718
No. evaluation points	260		360	
Bandwidth	0.15		0.20	
Observations	3027		3010	

Notes: Tests are based on 499 bootstrap samples.