Relaxing Instrument Exclusion with Common Confounders

Christian Tien (University of Cambridge)

Setup: Unobserved Common Confounding

Outcome $Y \in \mathcal{Y} \subseteq \mathbb{R}$, treatment $A \in \mathcal{A} \subseteq \mathbb{R}d_A$, instruments $Z \in \mathcal{Z} \subseteq \mathbb{R}d_Z$, proxies $W \in \mathcal{W} \subseteq \mathbb{R}d_W$, unobserved common confounders $U \in \mathcal{U} \subseteq \mathbb{R}d_U$. All results hold conditional on covariates. Notation:
$\Sigma_{XY} := \mathbb{E}[XY^\intercal] - \mathbb{E}[X]\,\mathbb{E}[Y]^\intercal$, $\Sigma_X = \Sigma_{XX}$, $\hat{W}_Z := \mathbb{E}_{\mathrm{L}}[W|Z] = Z\Sigma_Z^{-1}\Sigma_{ZW}$.

Assumption 1 (IV Common Confounding Model)**.** *1. SUTVA:* $Y = Y(A, Z)$.

2. Instruments

(a) Exclusion: $Y(a, z) = Y(a) \perp\!\!\!\perp Z \mid U$.

(b) Index sufficiency: For some $\tau \in L_2(Z)$*, where* $T := \tau(Z)$, $U \perp\!\!\!\perp Z \mid T$.

(c) Relevance (completeness): For any $g(A,T) \in L_2(A,T)$,

$$\mathbb{E}[g(A,T)|Z] = 0 \text{ only when } g(A,T) = 0. \quad (1)$$

3. Proxies

(a) Exclusion: $W \perp\!\!\!\perp Z \mid U$.

(b) Relevance (completeness): For any $g(U) \in L_2(U)$,

$$\mathbb{E}[g(U)|W] = 0 \text{ only when } g(U) = 0. \quad (2)$$

Linear Model Example

Equation	Exclusion	Relevance	
$Y = A\beta + W\upsilon_Y + U\gamma_Y + \varepsilon_Y,$	$\mathbb{E}[\varepsilon_Y Z] = \mathbf{0},$		(3)
$A = Z\zeta + W\upsilon_A + U\gamma_W + \varepsilon_A,$		$\operatorname{rank}\left(\mathbb{E}\left[A^T Z \mid \hat{W}_Z\right]\right) = d_A,$	(4)
$Z = U\gamma_Z + \varepsilon_Z,$		$\operatorname{rank}(\gamma_Z) = d_U < d_Z,$	(5)
$W = U\gamma_W + \varepsilon_W,$	$\mathbb{E}[\varepsilon_W^\intercal \varepsilon_Z] = \mathbf{0},$	$\operatorname{rank}(\gamma_W) = d_U \leq d_W.$	(6)

Idea: Identify a Valid Control from Observables

Quantity of interest: *Causal effect* of A on Y.	$J = \int Y(a)\pi(a)\mathrm{d}a$
A is endogenous (simultaneity, unobserved confounders).	$Y(a) \not\perp\!\!\!\perp A$
We want to use relevant instruments Z for A.	$A(z) \neq A$
Instruments NOT unconditionally excluded.	$Y(a) \not\perp\!\!\!\perp Z$
The unobserved *common confounders* U fully explain the association between Z and W.	$Z \perp\!\!\!\perp W \mid U$
Instruments would be excluded conditional on the common confounders U.	$Y(a) \perp\!\!\!\perp Z \mid U$

Lemma 1. *Assume* $W \perp\!\!\!\perp Z \mid U$ *(A1.3a), and for any* $g(U) \in L_2(U)$, $\mathbb{E}[g(U)|W] = 0$ *only when* $g(U) = 0$ *(A1.3b). Take any* $\tau \in L_2(Z)$*, where* $T := \tau(Z)$*, such that* $W \perp\!\!\!\perp Z \mid T$*. Then, also* $U \perp\!\!\!\perp Z \mid T$.

In words: If W and Z are independent conditional on $\tau(Z)$ (part of Z's variation), then so are U and Z conditional on $\tau(Z)$. Exclusion is restored. Identification ensues in with outcome model separability [Imbens and Newey, 2009] or first stage monotonicity [Newey and Powell, 2003]. Index sufficiency with fixed effects [Liu et al., 2021] similar in spirit.

Motivation/Application: Returns to Education

Data: National Longitudinal Survey of Youth 1997. $n =$ 1,983.

Y	Household net worth at 35: continuous variable, in USD
A	BA degree: 1 if individual i obtained a BA degree, 0 o.w.
Z	Pre-college test results: subject GPA, ASVAB percentile
W	Risky behaviour dummies: drinking, smoking, etc by age 17
U	Ability: Unmeasured intellectual capacity
ε_Y	Disturbance: Heterogeneous characteristics, and chance
X	Covariates: sex, college GPA, family variables, region, etc

Utility-maximiser chooses education with knowledge $\mathcal{I}$ about ε_Y and U:

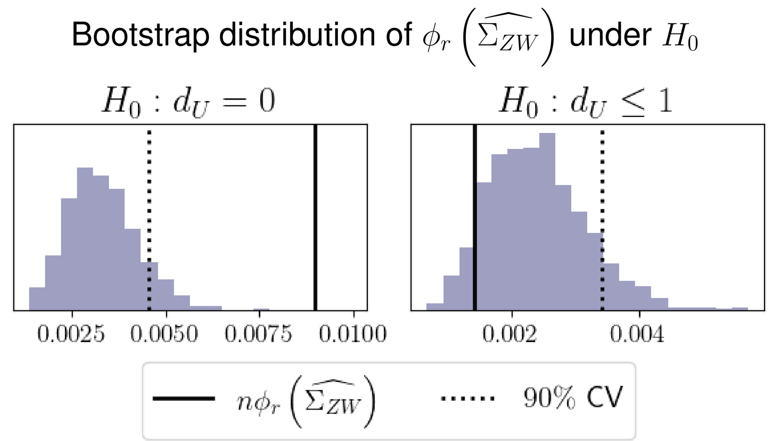
$$A = \underset{a \in \{0,1\}}{\arg\max} \left(\mathbb{E}\left[u(Y(a)) - c(a) | A = a, \mathcal{I}\right]\right), \quad (7)$$

Problem: Self-selection and unobserved confounding from ability ensue.

Test relevance of W and Z for U

For $r < \min\{d_Z, d_W\}$, use sum of r squared singular values $\phi_r(\Sigma_{ZW})$ to test

$$H_0 : d_U \leq r, \text{ vs } H_a : d_U > r. \quad (8)$$



Conclusion:

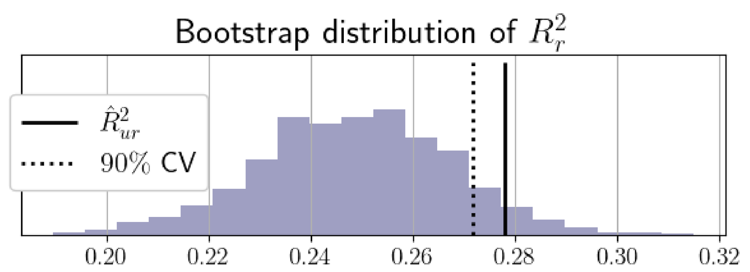
- Strong evidence for $d_U = 1$.
- W and Z relevant for U.

Test relevance of Z for A given T

Simple prediction test:

$H_0 : R^2_{ur} = R^2_r$, vs $R^2_{ur} > R^2_r$,

$$A = Z\tilde{\zeta} + X\tilde{\eta}_{A,ur} + \tilde{\varepsilon}_{A,ur} \implies R^2_{ur}$$
$$A = T\tilde{\gamma}_A + X\tilde{\eta}_{A,r} + \tilde{\varepsilon}_{A,r} \implies R^2_r$$



Conclusion: Z is relevant for A given T.

Exclusion of Z conditional on U

- Understand U through T to argue for/against exclusion (A1.2a).
- Here, T appears to hold a measure of general ability constant.
- Z probably excluded conditional on general ability and covariates.

Results

- Positive ability bias in OLS indicated by PL [Cui et al., 2020]
- Even larger negative selection bias in OLS indicated by IV
- ICC corrects for ability and selection bias
- About 50% larger SEs in ICC compared to IV

Estimates for β and linear parameter on T
(SE in parantheses)

	OLS	PL	IV	ICC
β	59.18 (9.12)	30.90 (10.40)	222.97 (34.74)	125.15 (52.93)
T		27.76 (4.82)		16.05 (7.37)

Intuition: Orthogonalisation wrt Unobservables

Create control function T

Decompose the covariance of Z and W as

$$\Sigma_{ZW} = \underbrace{\gamma_Z^\intercal \Sigma_U \gamma_W}_{\implies \operatorname{rank}(\Sigma_{ZW}) = d_U} := C_Z C_W^\intercal. \quad (9)$$

for some $C_Z \in \mathbb{R}^{d_Z \times d_U}$ and $C_W \in \mathbb{R}^{d_W \times d_U}$ s.t. a valid control function T is

$$T := Z\Sigma_Z^{-1}C_Z, \qquad \mathbb{E}_{\mathrm{L}}[W|Z] = TC_W^\intercal, \qquad \Sigma_{ZT} = C_Z, \Sigma_T = C_Z^\intercal \Sigma_Z^{-1} C_Z.$$

How does T help deconfound Z?

- $\mathbb{E}_{\mathrm{L}}[U|Z]$ is proportional to $\mathbb{E}_{\mathrm{L}}[W|Z]$, because $\operatorname{rank}(\gamma_W) = d_U$.
- By holding $\mathbb{E}_{\mathrm{L}}[W|Z]$ fixed via T, we are also holding $\mathbb{E}_{\mathrm{L}}[U|Z]$ fixed.
- All endogeneity in Z is from $\mathbb{E}_{\mathrm{L}}[U|Z]$, so fixed T restores exclusion.

E.g. for any $D_Z \in \mathbb{R}^{d_U \times d_A}$ s.t. $\operatorname{rank}\left(\left(I_{d_z} - \Sigma_T^{-1}\Sigma_{TZ}\right) D_Z\right) = d_A$ get deconfounded instrument $\tilde{Z}$ as

$$\tilde{Z} = Z \underbrace{\left(I_{d_z} - \Sigma_Z^{-1} C_Z \left(C_Z^\intercal \Sigma_Z^{-1} C_Z\right)^{-1} C_Z^\intercal\right)}_{:=M, \text{ note that } C_Z^\intercal M = \mathbf{0}} D_Z. \quad (10)$$

Consistent method of moments estimator

$$\beta = \Sigma_{\tilde{Z}A}^{-1}\Sigma_{\tilde{Z}Y} \implies \hat{\beta}_{MoM} = \left(D_Z^\intercal \hat{M}^\intercal Z^\intercal A\right)^{-1} \left(D_Z^\intercal \hat{M}^\intercal Z^\intercal Y\right) \quad (11)$$

Conclusion

- Using proxies to deconfound wrt common unobservables . . .
  - \+ can restore instrument exclusion when conditioning cannot, but
  - \- consumes more variation in the instruments than conditioning on observables, thus requiring rich relevance of instruments for treatment.
- Method most relevant with rich observational data, but intricate biases (like selection or simultaneity).

References

Yifan Cui, Hongming Pu, Xu Shi, Wang Miao, and Eric Tchetgen Tchetgen. Semiparametric proximal causal inference. *arXiv preprint arXiv:2011.08411*, 2020.

Guido W Imbens and Whitney K Newey. Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica*, 77(5):1481–1512, 2009.

Laura Liu, Alexandre Poirier, and Ji-Liang Shiu. Identification and estimation of average partial effects in semiparametric binary response panel models. *arXiv preprint arXiv:2105.12891*, 2021.

Whitney K Newey and James L Powell. Instrumental variable estimation of nonparametric models. *Econometrica*, 71(5):1565–1578, 2003.

arXiv:2301.02052 ct493@cam.ac.uk