Inference for Non-linear Treatment Effects with Control Function Methods

Zijian Guo

Department of Statistics
Rutgers University

February 25, 2021

Collaborators

Dylan Small

Sai Li

- Guo, Z., & Small, D. S. (2016). Control function instrumental variable estimation of nonlinear causal effect models. *Journal of Machine Learning Research*, 17(100), 1-35.
- Li, S., & Guo, Z. (2020). Causal Inference for Nonlinear Outcome Models with Possibly Invalid Instrumental Variables. *arXiv preprint arXiv:2010.09922.*

Overview of talk

Endogeneity and Instrumental Variable

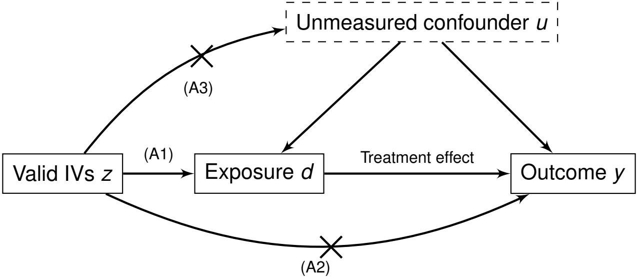


Figure: IV assumptions (A1)-(A3).

(A1) association with the treatment;

(A2) no direct effect on the outcome;

(A3) ignorability.

Overview of talk

1. Endogeneity and Instrumental Variable
2. Control Function and TSLS
3. Control Function with Possibly Invalid IVs

Models

Outcome model

$$y_i = \beta_0 + d_i\beta_1 + d_i^2\beta_2 + x_i^{\intercal}\psi + u_i, \quad \text{for} \quad 1 \leq i \leq n$$

Treatment model

$$d_i = z_i\gamma_1 + z_i^2\gamma_2 + x_i^{\intercal}\phi + v_i \quad \text{for} \quad 1 \leq i \leq n$$

- baseline covariate x_i
- u_i is correlated with v_i and hence d_i
- The result can be extended to **known** h

$$y_i = h(d_i) + x_i^{\intercal}\psi + u_i$$

IV and TSLS

1. Predict d by $\widehat{d}$

$$\mathrm{lm}(d \sim z + z^2 + x)$$

   Predict d^2 by $\widehat{d^2}$

$$\mathrm{lm}(d^2 \sim z + z^2 + x)$$

2. Run a second stage regression

$$\mathrm{lm}(y \sim \widehat{d} + \widehat{d^2} + x)$$

Estimate β_1 and β_2 by coefficients in front of $\widehat{d}$ and $\widehat{d^2}$.

IV and Control Function

1. Predict d by $\widehat{d}$

$$\mathrm{lm}(d \sim z + z^2 + x)$$

   and obtain the residual $e_1 = d - \widehat{d}$.

2. Run a second stage regression

$$\mathrm{lm}(Y \sim d + d^2 + x + e_1)$$

Estimate β_1 and β_2 by coefficients in front of d and d^2.

IV and Control Function

1. Predict d by $\widehat{d}$

$$\mathrm{lm}(d \sim z + z^2 + x)$$

   and obtain the residual $e_1 = d - \widehat{d}$.

2. Run a second stage regression

$$\mathrm{lm}(Y \sim d + d^2 + x + e_1)$$

Estimate β_1 and β_2 by coefficients in front of d and d^2.

- e_1 is a surrogate for part of the unmeasured confounder in d.
- Two Stage Residual Inclusion

Intuition and Assumption

If v is known, then

$$\begin{aligned}\mathbb{E}(y_i \mid d_i, x_i, \mathbf{v_i}) &= \beta_0 + \beta_1 d_i + \beta_2 d_i^2 + x_i^{\intercal}\psi + \mathbb{E}(u_i \mid d_i, x_i, v_i)\\ &= \beta_0 + \beta_1 d_i + \beta_2 d_i^2 + x_i^{\intercal}\psi + \mathbb{E}(u_i \mid z_i, x_i, v_i)\\ &= \beta_0 + \beta_1 d_i + \beta_2 d_i^2 + x_i^{\intercal}\psi + \mathbb{E}(u_i \mid v_i)\\ &= \beta_0 + \beta_1 d_i + \beta_2 d_i^2 + x_i^{\intercal}\psi + \rho v_i\end{aligned}$$

Assumptions

1. (u_i, v_i) are independent of z_i, x_i
2. $\mathbb{E}(u_i \mid v_i) = \rho v_i$

Intuition and Assumption

If v is known, then

$$\begin{aligned}\mathbb{E}(y_i \mid d_i, x_i, \mathbf{v_i}) &= \beta_0 + \beta_1 d_i + \beta_2 d_i^2 + x_i^\intercal \psi + \mathbb{E}(u_i \mid d_i, x_i, v_i) \\ &= \beta_0 + \beta_1 d_i + \beta_2 d_i^2 + x_i^\intercal \psi + \mathbb{E}(u_i \mid z_i, x_i, v_i) \\ &= \beta_0 + \beta_1 d_i + \beta_2 d_i^2 + x_i^\intercal \psi + \mathbb{E}(u_i \mid v_i) \\ &= \beta_0 + \beta_1 d_i + \beta_2 d_i^2 + x_i^\intercal \psi + \rho v_i\end{aligned}$$

Assumptions

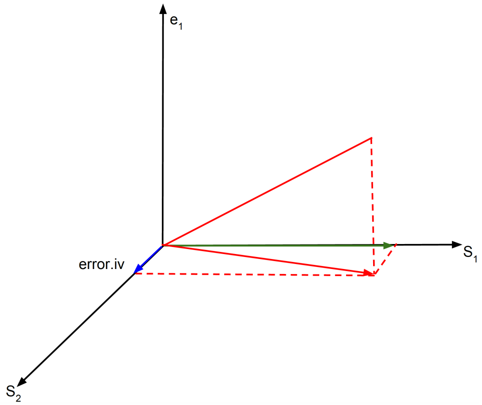
1. (u_i, v_i) are independent of z_i, x_i
2. $\mathbb{E}(u_i \mid v_i) = \rho v_i$

Imbens, W.G and Wooldridge, M.J. *Control Function and Related Methods*, Lecture Notes on course "What's New in Econometrics ", NBER (2007).

- If the outcome model is linear in d, then TSLS=CF.

CF: Augmented TSLS

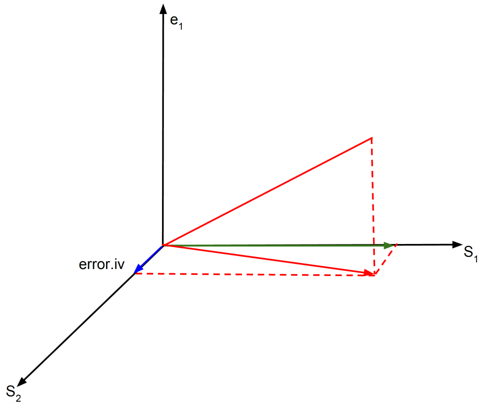
- Define $\widetilde{d^2}$ as the residual of the regression $d^2 \sim e_1$.
- Define $error.iv = \text{resid}(\widetilde{d^2} \sim x + z + z^2)$.
- $S_1 = span\{1, x, z, z^2\}$ and $S_2 = span\{error.iv\}$.



CF: Augmented TSLS

Theorem 1

The Control Function Estimator with Instruments x, z, z^2 is the same with TSLS with Instruments x, z, z^2 and $error.iv$.
If $error.iv$ is a valid instrument, CF is more efficient than 2SLS.



Validity and Hausman Test

Define $V_0 = (1, x, z, z^2)$ and $V = (1, x, z, z^2, error.iv)$ and $W = (1, x, d, d^2)$. Under the conditional homoskedasticity, we define

$$\widehat{\eta}_0 = (W^\intercal P_0 W)^{-1} W P_0 Y \quad \text{with} \quad P_0 = V_0 (V_0^\intercal V_0)^{-1} V_0^\intercal$$

$$\widehat{\eta} = (W^\intercal P W)^{-1} W P Y \quad \text{with} \quad P = V (V^\intercal V)^{-1} V^\intercal$$

$$C = \frac{\hat{u}^\intercal P \hat{u} - \widehat{u}_0^\intercal P_0 \widehat{u}_0}{\hat{\sigma}^2} \quad \text{is asymptotically} \quad \chi^2(1)$$

where

$$\hat{u} = y - W\widehat{\eta},\ \widehat{u}_0 = y - W\widehat{\eta}_0,\ \hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{n}.$$

Hayashi,F. *Econometrics*, Princeton University Press. (2000)

Pretest Estimator

Define the p-value $p = P(\chi^2(1) \geq C)$. **The Level α Pretest Estimator** is defined as

$$\begin{cases} CF & \text{if } p > \alpha \\ TSLS & \text{if } p \leq \alpha \end{cases}$$

Simulation 1

$$y = 1 + x + 10d + 10d^2 + u$$
$$d = 1 + \frac{1}{8}x + \frac{1}{3}z + \frac{1}{8}z^2 + v$$

where $x \sim N(0, 10^2)$, $z \sim N(0, 3^2)$ and
$\begin{pmatrix} u \\ v \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 100 & 31 \\ 31 & 10 \end{pmatrix}\right]$.

Simulation 1

	WMSE		NMSE	
	CF	Pretest	CF	Pretest
β_0	0.79	0.83	0.29	0.40
β_1	0.12	0.13	0.03	0.28
β_2	0.04	0.04	0.01	0.28
β_3	0.01	0.01	0.001	0.29

Table: Proportion of Winsorized MSE (WMSE) and Non-winsorized MSE(NMSE) of the estimators, with WMSE/MSE of TSLS as basis, Sample size 10,000 and simulation time is 10,000, pvalue>0.05

Simulation 2

$$y = d + 0.2d^2 + w + u$$
$$d = -1 + 0.2z + 0.3z^2 + v$$
$$w = 0.5v^2 + N(0,1)$$

where $\begin{pmatrix} u \\ v \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right]$ and $z \sim N(0,1)$.

$$\mathbb{E}(w_i + u_i \mid v_i) = 0.5v_i^2 \neq \rho v_i$$

Simulation 2

	Bias of Sample Mean		
	TSLS	CF	Pretest
β_1	$4e^{-6}$	-0.128	$4e^{-6}$
β_2	$-1.6e^{-4}$	0.559	$-1.6e^{-4}$

Table: Proportion of Bias of Sample Mean of the estimators to the true value, Sample size 10,000 and simulation time is 10,000, pvalue>0.05

Simulation 2

	WMSE		NMSE	
	CF	Pretest	CF	Pretest
β_1	6.91	1	6.31	1
β_2	10.14	1	9.24	1

Table: Proportion of Winsorized MSE (WMSE) and Non-winsorized MSE(NMSE) of the estimators, with WMSE/MSE of TSLS as basis.

Take Home Message

1. Control function = TSLS with an augmented set of IVs
2. Pretest estimator: combining CF and TSLS

Guo, Z., & Small, D. S. (2016). Control function instrumental variable estimation of nonlinear causal effect models. *Journal of Machine Learning Research*, 17(100), 1-35.

Code is available at `https://github.com/zijguo/Control-function`.

Overview of talk

1. Endogeneity and Instrumental Variable
2. Control Function and TSLS
3. Control Function with Possibly Invalid IVs

Binary Outcome and Invalid IVs

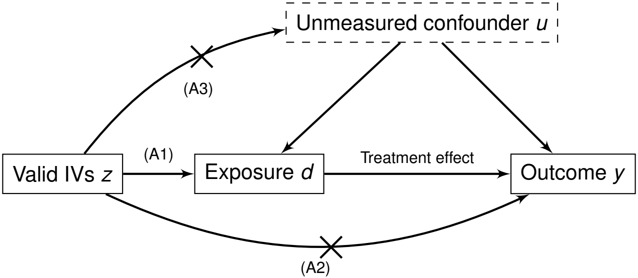


Figure: IV assumptions (A1)-(A3).

Binary Outcome+ Violation of (A2) and (A3).

Model Set-up

Define $w_i = (z_i^\intercal, x_i^\intercal)^\intercal$. Potential outcome model

$$\mathbb{E}[y_i^{(d)} | w_i = w, u_i = u] = q\left(d\beta + w^\intercal \kappa, u\right),$$

where $\kappa = (\kappa_z^\intercal, \kappa_x^\intercal)^\intercal$ and $q : \mathbb{R}^2 \to \mathbb{R}$ is a possibly unknown function.

Define $w_i = (z_i^\intercal, x_i^\intercal)^\intercal$. Potential outcome model

$$\mathbb{E}[y_i^{(d)} | w_i = w, u_i = u] = q\left(d\beta + w^\intercal \kappa, u\right),$$

where $\kappa = (\kappa_z^\intercal, \kappa_x^\intercal)^\intercal$ and $q : \mathbb{R}^2 \to \mathbb{R}$ is a possibly unknown function.

- Logistic

$$q(d\beta + w^\intercal \kappa, u) = \frac{\exp(d\beta + w^\intercal \kappa + u)}{1 + \exp(d\beta + w^\intercal \kappa + u)}$$

Define $w_i = (z_i^\intercal, x_i^\intercal)^\intercal$. Potential outcome model

$$\mathbb{E}[y_i^{(d)} | w_i = w, u_i = u] = q\left(d\beta + w^\intercal \kappa, u\right),$$

where $\kappa = (\kappa_z^\intercal, \kappa_x^\intercal)^\intercal$ and $q : \mathbb{R}^2 \to \mathbb{R}$ is a possibly unknown function.

- Logistic

$$q(d\beta + w^\intercal \kappa, u) = \frac{\exp(d\beta + w^\intercal \kappa + u)}{1 + \exp(d\beta + w^\intercal \kappa + u)}$$

- Probit (standard normal u)

$$q(d\beta + w^\intercal \kappa, u) = \mathbf{1}(d\beta + w^\intercal \kappa + u > 0)$$

Define $w_i = (z_i^\intercal, x_i^\intercal)^\intercal$. Potential outcome model

$$\mathbb{E}[y_i^{(d)} | w_i = w, u_i = u] = q\left(d\beta + w^\intercal \kappa, u\right),$$

where $\kappa = (\kappa_z^\intercal, \kappa_x^\intercal)^\intercal$ and $q : \mathbb{R}^2 \to \mathbb{R}$ is a possibly unknown function.

- Logistic

$$q(d\beta + w^\intercal \kappa, u) = \frac{\exp(d\beta + w^\intercal \kappa + u)}{1 + \exp(d\beta + w^\intercal \kappa + u)}$$

- Probit (standard normal u)

$$q(d\beta + w^\intercal \kappa, u) = \mathbf{1}(d\beta + w^\intercal \kappa + u > 0)$$

- Continuous outcome models

$$q\left(d\beta + w^\intercal \kappa, u\right) = \left(d\beta + w^\intercal \kappa\right) \cdot u$$

$$\mathbb{E}[y_i^{(d)} | w_i = w, u_i = u] = q\,(d\beta + w^{\intercal}\kappa, u)$$

- q can be unknown.
- u_i and d_i are correlated.
- $\kappa_z \neq 0$ indicates a direct effect!

$$\mathbb{E}[y_i^{(d)}|w_i = w, u_i = u] = q\left(d\beta + w^{\intercal}\kappa, u\right)$$

- q can be unknown.
- u_i and d_i are correlated.
- $\kappa_z \neq 0$ indicates a direct effect!
- The target causal estimand is CATE

$$\mathrm{CATE}(d, d'|w) := \mathbb{E}\left[y_i^{(d)} - y_i^{(d')}|w_i = w\right],$$

where $d \in \mathbb{R}$ and $d' \in \mathbb{R}$ and $w \in \mathbb{R}^p$.

Potential outcome model and consistency imply

$$\begin{aligned}\mathbb{E}[y_i|d_i = d, w_i = w, u_i = u] &= q\,(d\beta + w^\intercal \kappa, u) \\ &= \frac{\exp(d\beta + w^\intercal \kappa + u)}{1 + \exp(d\beta + w^\intercal \kappa + u)}\end{aligned}$$

Continuous treatment model

$$d_i = w_i^\intercal \gamma + v_i, \quad \mathbb{E}[v_i|w_i] = 0,$$

where $\gamma = (\gamma_z^\intercal, \gamma_x^\intercal)^\intercal$ and v_i is the residual term.

Inference for β.

Existing CF Methods

Blundell, R. W. and J. L. Powell (2004). Endogeneity in semiparametric binary response models. The Review of Economic Studies 71(3), 655–679.

Rothe, C. (2009). Semiparametric estimation of binary response models with endogenous regressors. Journal of Econometrics 153(1), 51–64.

Literature results

Classical CF Assumptions

(A1) $\|\gamma_z\|_2 \geq \tau_0 > 0$ for some $\tau_0 > 0$;

(A2) $\kappa_z = 0$;

(A3) $f_u(u_i|w_i, v_i) = f_u(u_i|v_i)$ where $w_i = (z_i^\intercal, x_i^\intercal)^\intercal$.

If w_i is independent of (u_i, v_i), (A3) holds.

Literature results

Classical CF Assumptions

(A1) $\|\gamma_z\|_2 \geq \tau_0 > 0$ for some $\tau_0 > 0$;

(A2) $\kappa_z = 0$;

(A3) $f_u(u_i|w_i, v_i) = f_u(u_i|v_i)$ where $w_i = (z_i^\intercal, x_i^\intercal)^\intercal$.

If w_i is independent of (u_i, v_i), (A3) holds.

(A2) and (A3) imply

$$\mathbb{E}[y_i|d_i, w_i, v_i] = \int q(d_i\beta + w_i^\intercal \kappa, u_i) f_u(u_i|v_i) du_i = g_0\left(d_i\beta + x_i^\intercal \kappa_x, v_i\right)$$

1. Double index model.
2. The literature is about inference for β.

New Identifiability Conditions

Dimension reduction condition:

$$f_u(u_i|w_i, v_i) = f_u(u_i|w_i^{\intercal}\eta, v_i) \quad \text{for some } \eta \in \mathbb{R}^{p\times q}. \tag{1}$$

- $\eta \neq 0$: non-parametric violation of (A3) .
- Focus on $q = 1$

Dimension reduction condition:

$$f_u(u_i|w_i, v_i) = f_u(u_i|w_i^{\intercal}\eta, v_i) \quad \text{for some } \eta \in \mathbb{R}^{p\times q}. \tag{1}$$

- $\eta \neq 0$: non-parametric violation of (A3) .
- Focus on $q = 1$

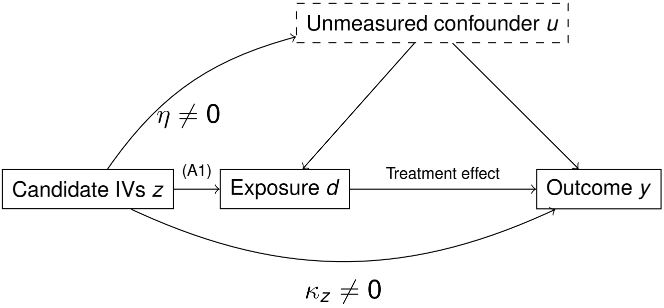
Majority rule: more than half of the relevant IVs are valid.

- set of relevant IVs

$$\mathcal{S} = \{1 \leq j \leq p_z : \gamma_j \neq 0\}.$$

- set of valid IVs

$$\mathcal{V} = \{j \in \mathcal{S} : (\kappa_z)_j = (\eta_z)_j = 0\}.$$



$$
\begin{aligned}
\mathbb{E}[y_i|d_i, w_i, {\color{red}v_i}] &= \int q(d_i\beta + w_i^{\intercal}\kappa, u_i) f_u(u_i|w_i^{\intercal}\eta, v_i) du_i \\
&= {\color{red} g^*\left(d_i\beta + w_i^{\intercal}\kappa, w_i^{\intercal}\eta, v_i\right)}
\end{aligned}
$$

1. We allow $\kappa_z \neq 0$ and $\eta \neq 0$
2. In comparison to ${\color{blue} g_0\left(d_i\beta + x_i^{\intercal}\kappa_x, v_i\right)}$

Identifiability Strategy

Step 1: Reduced Form Estimators

Expressed in the matrix form,

$$\mathbb{E}[y_i|d_i, w_i, v_i] = g^*((d_i, w_i^{\intercal})B^*, v_i) \quad \text{with} \quad B^* = \begin{pmatrix} \beta & 0 \\ \kappa & \eta \end{pmatrix} \in \mathbb{R}^{(p+1)\times 2}.$$

We plugin $d_i = w_i^{\intercal}\gamma + v_i$ and obtain

$$\mathbb{E}[y_i|w_i, v_i] = \mathbb{E}[y_i|w_i^{\intercal}\Theta^*, v_i] \quad \text{with} \quad \Theta^* = \begin{pmatrix} \beta\gamma + \kappa & \eta \end{pmatrix} \in \mathbb{R}^{p\times 2}.$$

Estimate Θ^* by standard dimension reduction methods.

Step 2: Apply Majority Rule

Identify Θ as a linear transformation of Θ^* :

$$\Theta^* = \begin{pmatrix} \beta\gamma + \kappa & \eta \end{pmatrix} \in \mathbb{R}^{p\times 2}.$$

Define

$$b_m = \mathrm{Median}(\{\Theta_{j,m}/\gamma_j\}_{j\in\mathcal{S}}) \quad \text{for} \quad 1 \leq m \leq 2.$$

where $\mathcal{S}$ denotes the set of relevant IV. We identify B as

$$B = \begin{pmatrix} b_1 & b_2 \\ \Theta_{\cdot,1} - b_1\gamma & \Theta_{\cdot,2} - b_2\gamma \end{pmatrix} \tag{2}$$

Construct B such that

$$\mathbb{E}\left[y_i^{(d)} | w_i = w, v_i = v\right] = g\left((d, w^\intercal)B, v\right)$$

Step 3: partial mean

$$\mathrm{CATE}(d, d'|w) := \mathbb{E}\left[y_i^{(d)} - y_i^{(d')}|w_i = w\right],$$

Identify $\mathbb{E}\left[y_i^{(d)}|w_i = w\right]$ by

$$\int \mathbb{E}\left[y_i^{(d)}|w_i = w, v_i = v\right] f_v(v)dv$$

Average with respect to v_i : $\frac{1}{n}\sum_{i=1}^{n} g((d, w^{\intercal})B, v_i)$.

Informal Theoretical Results

Under regularity conditions,

$$\frac{n}{\sqrt{\mathrm{V}_{\mathrm{CATE}}}}\left(\widehat{\mathrm{CATE}}(d,d'|w)-\mathrm{CATE}(d,d'|w)\right)\to N(0,1)$$

and

$$\mathbf{P}\left(c_0/\sqrt{nh^2}\le\sqrt{\mathrm{V}_{\mathrm{CATE}}}/n\le C_0/\sqrt{nh^2}\right)\ge 1-n^{-c}.$$

1. Confidence interval is constructed by bootstrap.
2. Similar to two-dimension non-parametric function!
3. Inference for CATE is much more challenging than inference for β.

Real Data: Factor IV

Mouse data set (Bush and Moore 2012).

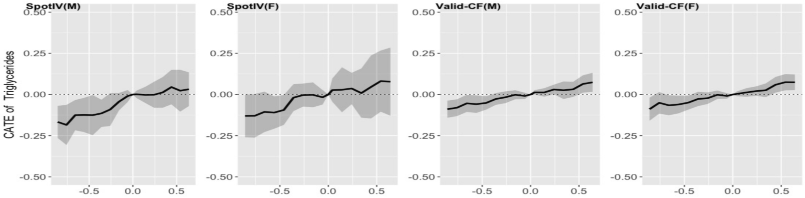
- 10,346 polymorphic genetic markers and 1,269 sample
- outcome: (pre) diabetic v.s. normal
- exposures: HDL, LDL, Triglycerides
- a large number of polymorphic markers
- the high correlation among some polymorphic markers.

Factor IV

1. Select polymorphic markers which have “not-too-small” marginal associations with HDL
2. Run PCA and use leading PC as the candidate IVs.
3. HDL (24 IVs); LDL (18 IVs); Triglycerides (14 IVs)

Real Data Results

The constructed 95% CIs for $\text{CATE}(d, 0|w_M)$ and $\text{CATE}(d, 0|w_F)$ with Triglycerides exposures at different levels of d.



Take Home Message

1. New ways to model invalid IVs.
2. New identifiability conditions for control function.
3. Confidence interval construction for the treatment effect.

Li, S., & Guo, Z. (2020). Causal Inference for Nonlinear Outcome Models with Possibly Invalid Instrumental Variables. *arXiv preprint arXiv:2010.09922.*

Code is available at `https://github.com/saili0103/SpotIV`.

Acknowledgement

Special thanks to

- Collaborators: Dylan Small, Tony Cai, Hyunseung Kang and Sai Li
- Package Contributors: Taehyeon Koo, Wei Yuan and Yunjiao Bai

Acknowledgement to NSF and NIH for fundings.

Thank you!