UNIVERSITY OF CAMBRIDGE

Sensitivity Analysis for Instrumental Variables

Tobias Freidling, Qingyuan Zhao

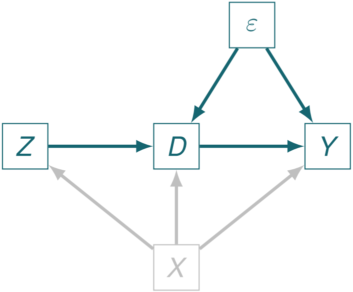
DPMMS and Cantab Capital Institute for the Mathematics of Information,
University of Cambridge, UK

Instrumental Variables

Goal: Estimation of and confidence interval for the causal effect $D \to Y$ under confounding ε via instrumental variables Z

Example:

- Y: wage
- D: education
- ε: ability
- Z: college proximity

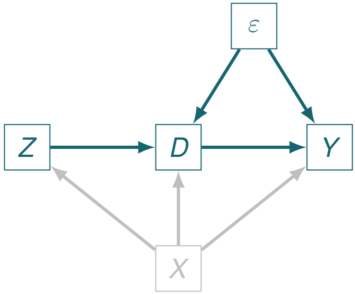


Instrumental Variables

Goal: Estimation of and confidence interval for the causal effect $D \to Y$ under confounding ε via instrumental variables Z

Example:

- Y: wage
- D: education
- ε: ability
- Z: college proximity



Instrumental Variables

Definition

Assume i.i.d. data $Y \in \mathbb{R}^n, D \in \mathbb{R}^{n \times p}, Z \in \mathbb{R}^{n \times k}$.

$$Y = D\beta + \varepsilon_Y, \qquad D = Z\Gamma + \varepsilon_D, \qquad \varepsilon = [\varepsilon_Y : \varepsilon_D],$$
$$\mathbb{E}[\varepsilon_i | Z_i] = 0, \ \mathrm{Var}(\varepsilon_i | Z_i) = \Sigma$$

The parameters are $\beta \in \mathbb{R}^p$, $\Gamma \in \mathbb{R}^{k \times p}$ and $\Sigma \in \mathbb{R}^{(p+1) \times (p+1)}$.

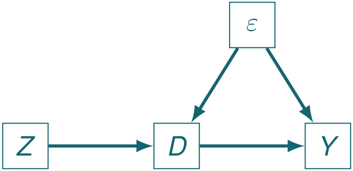
Assumptions

(A1) $\mathbb{E}[Z^T D]$ has rank p.

(A2) $\mathbb{E}[Z^T Z]$ has rank k.

(A3) $\mathbb{E}[Z^T \varepsilon] = 0$.

(I) No other causal pathways.



Instrumental Variables

Definition

Assume i.i.d. data $Y \in \mathbb{R}^n, D \in \mathbb{R}^{n\times p}, Z \in \mathbb{R}^{n\times k}$.

$$Y = D\beta + \varepsilon_Y, \qquad D = Z\Gamma + \varepsilon_D, \qquad \varepsilon = [\varepsilon_Y : \varepsilon_D],$$
$$\mathbb{E}[\varepsilon_i | Z_i] = 0,\ \mathrm{Var}(\varepsilon_i | Z_i) = \Sigma$$

The parameters are $\beta \in \mathbb{R}^p$, $\Gamma \in \mathbb{R}^{k\times p}$ and $\Sigma \in \mathbb{R}^{(p+1)\times(p+1)}$.

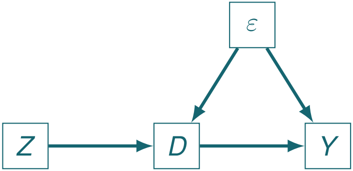
Assumptions

(A1) $\mathbb{E}[Z^T D]$ has rank p.

(A2) $\mathbb{E}[Z^T Z]$ has rank k.

(A3) $\mathbb{E}[Z^T \varepsilon] = 0$.

(I) No other causal pathways.



Sensitivity Model

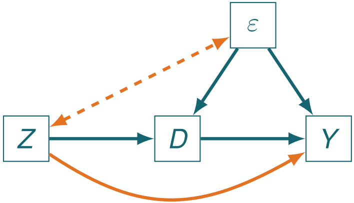
(A1) is widely researched and (A2) is easily satisified.

Definition (Linear IV Sensitivity Model)

Assume i.i.d. data $Y \in \mathbb{R}^n, D \in \mathbb{R}^{n \times p}, Z \in \mathbb{R}^{n \times k}$ and let $\Delta \subset \mathbb{R}^k$ bounded.

$$Y = D\beta + Z\delta + \varepsilon_Y, \qquad D = Z\Gamma + \varepsilon_D, \qquad \varepsilon = [\varepsilon_Y : \varepsilon_D],$$
$$\mathbb{E}[\varepsilon_i | Z_i] = 0, \ \mathrm{Var}(\varepsilon_i | Z_i) = \Sigma$$

The parameters are $\beta \in \mathbb{R}^p$, $\Gamma \in \mathbb{R}^{k \times p}$, $\delta \in \Delta$ and $\Sigma \in \mathbb{R}^{(p+1)\times(p+1)}$.



Sensitivity Region

The sensitivity model is overparametrised and thus β is only partially identified. Abbreviate $\Pi = (\Gamma, \Sigma)$ and denote the distribution of the model $\mathcal{F}_{\beta,\Pi,\delta}$. Two sets of parameters (β, Π, δ) and (β', Π', δ') are observationally equivalent if the corresponding distributions are equal, $\mathcal{F}_{\beta,\Pi,\delta} \simeq \mathcal{F}_{\beta',\Pi',\delta'}$.

Definition (Sensitivity Interval/Region)

Any $1-\alpha$ sensitivity region S_Δ for the sensitivity set Δ must satisfy

$$\inf_{\mathcal{F}_{\beta,\Pi,\delta} \simeq \mathcal{F}_{\beta_0,\Pi_0,\delta_0}} \mathbb{P}_{\beta_0,\Pi_0,\delta_0}(\beta \in S_\Delta) \geq 1-\alpha, \quad \forall\, \beta_0, \Pi_0, \delta_0 \in \Delta.$$

Sensitivity Region

The sensitivity model is overparametrised and thus β is only partially identified. Abbreviate $\Pi = (\Gamma, \Sigma)$ and denote the distribution of the model $\mathcal{F}_{\beta,\Pi,\delta}$. Two sets of parameters (β, Π, δ) and (β', Π', δ') are observationally equivalent if the corresponding distributions are equal, $\mathcal{F}_{\beta,\Pi,\delta} \simeq \mathcal{F}_{\beta',\Pi',\delta'}$.

Definition (Sensitivity Interval/Region)

Any $1-\alpha$ sensitivity region S_Δ for the sensitivity set Δ must satisfy

$$\inf_{\mathcal{F}_{\beta,\Pi,\delta} \simeq \mathcal{F}_{\beta_0,\Pi_0,\delta_0}} \mathbb{P}_{\beta_0,\Pi_0,\delta_0}\left(\beta \in S_\Delta\right) \geq 1-\alpha, \quad \forall\, \beta_0, \Pi_0, \delta_0 \in \Delta.$$

Union Method

For any fixed $\delta \in \Delta$, we can apply standard IV theory by replacing Y with $Y - Z\delta$: estimation of $\beta(\delta)$ and asymptotic confidence interval $I^{(\delta)} = [L^{(\delta)}, U^{(\delta)}]$. If the $I^{(\delta)}$ are congruent,

$$\bigcup_{\delta \in \Delta} I^{(\delta)} \subset \left[\inf_{\delta \in \Delta} L^{(\delta)}, \sup_{\delta \in \Delta} U^{(\delta)} \right]$$

is indeed a sensitivity interval. Hence, we must solve an optimisation problem.

Inversion of Tests

Suppose we can test H_0: $\beta = \beta^*$ against H_1: $\beta \neq \beta^*$ at level α for any β^*, i.e. under H_0

$$\mathbb{P}_{\beta^*,\Pi,\delta}(D \in A(\beta^*)) \geq 1 - \alpha, \quad \forall\, \Pi, \delta \in \Delta,$$

where D denotes the data and $A(\beta^*)$ is region of the test.

Define $S_\Delta = \{\beta^* \colon D \in A(\beta^*)\}$, then

$$\beta^* \in S_\Delta \Leftrightarrow D \in A(\beta^*), \qquad \mathbb{P}_{\beta,\Pi,\delta}(\beta \in S_\Delta) \geq 1 - \alpha, \quad \forall\, \beta, \Pi, \delta \in \Delta.$$

Take infimum over the observationally equivalent distributions:

$$\inf_{\mathcal{F}_{\beta,\Pi,\delta} \simeq \mathcal{F}_{\beta_0,\Pi_0,\delta_0}} \mathbb{P}_{\beta,\Pi,\delta}(\beta \in S_\Delta) = \inf_{\mathcal{F}_{\beta,\Pi,\delta} \simeq \mathcal{F}_{\beta_0,\Pi_0,\delta_0}} \mathbb{P}_{\beta_0,\Pi_0,\delta_0}(\beta \in S_\Delta)$$
$$\geq 1 - \alpha, \qquad \forall\, \beta_0, \Pi_0, \delta_0 \in \Delta.$$

Inversion of Tests

Suppose we can test H_0: $\beta = \beta^*$ against H_1: $\beta \neq \beta^*$ at level α for any β^*, i.e. under H_0

$$\mathbb{P}_{\beta^*,\Pi,\delta}(D \in A(\beta^*)) \geq 1 - \alpha, \quad \forall\, \Pi, \delta \in \Delta,$$

where D denotes the data and $A(\beta^*)$ is region of the test.
Define $S_\Delta = \{\beta^* \colon D \in A(\beta^*)\}$, then

$$\beta^* \in S_\Delta \Leftrightarrow D \in A(\beta^*), \qquad \mathbb{P}_{\beta,\Pi,\delta}(\beta \in S_\Delta) \geq 1 - \alpha, \quad \forall\, \beta, \Pi, \delta \in \Delta.$$

Take infimum over the observationally equivalent distributions:

$$\inf_{\mathcal{F}_{\beta,\Pi,\delta} \simeq \mathcal{F}_{\beta_0,\Pi_0,\delta_0}} \mathbb{P}_{\beta,\Pi,\delta}(\beta \in S_\Delta) = \inf_{\mathcal{F}_{\beta,\Pi,\delta} \simeq \mathcal{F}_{\beta_0,\Pi_0,\delta_0}} \mathbb{P}_{\beta_0,\Pi_0,\delta_0}(\beta \in S_\Delta)$$
$$\geq 1 - \alpha, \qquad \forall\, \beta_0, \Pi_0, \delta_0 \in \Delta.$$

Inversion of Tests

Suppose we can test H_0: $\beta = \beta^*$ against H_1: $\beta \neq \beta^*$ at level α for any β^*, i.e. under H_0

$$\mathbb{P}_{\beta^*,\Pi,\delta}(D \in A(\beta^*)) \geq 1 - \alpha, \quad \forall\, \Pi, \delta \in \Delta,$$

where D denotes the data and $A(\beta^*)$ is region of the test.
Define $S_\Delta = \{\beta^* \colon D \in A(\beta^*)\}$, then

$$\beta^* \in S_\Delta \Leftrightarrow D \in A(\beta^*), \qquad \mathbb{P}_{\beta,\Pi,\delta}(\beta \in S_\Delta) \geq 1 - \alpha, \quad \forall\, \beta, \Pi, \delta \in \Delta.$$

Take infimum over the observationally equivalent distributions:

$$\inf_{\mathcal{F}_{\beta,\Pi,\delta} \simeq \mathcal{F}_{\beta_0,\Pi_0,\delta_0}} \mathbb{P}_{\beta,\Pi,\delta}\left(\beta \in S_\Delta\right) = \inf_{\mathcal{F}_{\beta,\Pi,\delta} \simeq \mathcal{F}_{\beta_0,\Pi_0,\delta_0}} \mathbb{P}_{\beta_0,\Pi_0,\delta_0}\left(\beta \in S_\Delta\right)$$
$$\geq 1 - \alpha, \qquad \forall\, \beta_0, \Pi_0, \delta_0 \in \Delta.$$

Likelihood Ratio Statistic

We consider an i.i.d. data sample and a parametric statistical model $\{\mathbb{P}_\theta : \theta \in \Theta\}$ with log-likelihood $\ell_n(\theta)$. The maximum likelihood estimator for $\Theta^* \subset \Theta$ is defined as

$$\hat{\theta}_{\Theta^*} = \operatorname*{argmax}_{\theta \in \Theta^*} \ell_n(\theta).$$

Let $\Theta_0 \subset \Theta_1 \subset \Theta$ be nested models. To test $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1 \setminus \Theta_0$, we use the likelihood ratio statistic

$$\lambda_n = 2\left(\sup_{\theta \in \Theta_1} \ell_n(\theta) - \sup_{\theta \in \Theta_0} \ell_n(\theta)\right) = 2\left(\ell_n(\hat{\theta}_{\Theta_1}) - \ell_n(\hat{\theta}_{\Theta_0})\right).$$

Constrained Statistical Inference

Under regularity assumptions, if Θ_0 and Θ_1 are linear spaces, then $\lambda_n \xrightarrow{D} \chi^2_d$ as $n \to \infty$, where $d = \dim(\Theta_1) - \dim(\Theta_0)$.

Proposition (Silvapulle and Sen (2005))

If Θ_0 is "nice", e.g. defined by polynomial inequalities and equations, and $\Theta_1 = \Theta$, then

$$\lambda_n \xrightarrow{D} \sum_{i=0}^{m(\theta_0)} w_{m(\theta_0)-i}(m(\theta_0), V(\theta_0))\, \chi^2_{r(\theta_0)+i}, \quad \text{as } n \to \infty,$$

where θ_0 is the true value and $\sum_{i=0}^{m} w_{m-i} = 1$ for $w_{m-i} > 0$.

The asymptotic distribution depends on the unknown true value θ_0. Hence, we use the least favourable null.

Constrained Statistical Inference

Under regularity assumptions, if Θ_0 and Θ_1 are linear spaces, then $\lambda_n \xrightarrow{\text{D}} \chi^2_d$ as $n \to \infty$, where $d = \dim(\Theta_1) - \dim(\Theta_0)$.

Proposition (Silvapulle and Sen (2005))

If Θ_0 is "nice", e.g. defined by polynomial inequalities and equations, and $\Theta_1 = \Theta$, then

$$\lambda_n \xrightarrow{D} \sum_{i=0}^{m(\theta_0)} w_{m(\theta_0)-i}(m(\theta_0), V(\theta_0))\, \chi^2_{r(\theta_0)+i}, \quad \text{as } n \to \infty,$$

where θ_0 is the true value and $\sum_{i=0}^{m} w_{m-i} = 1$ for $w_{m-i} > 0$.

The asymptotic distribution depends on the unknown true value θ_0. Hence, we use the least favourable null.

Application to Sensitivity Analysis

Original model

$$Y = D\beta + Z\delta + \varepsilon_Y,$$
$$D = Z\Gamma + \varepsilon_D.$$

Reduced model

$$Y = Z\rho + \varepsilon_Y,$$
$$D = Z\Gamma + \varepsilon_D.$$

The original and reduced model are linked via $\rho = \delta + \Gamma\beta$. Hence, the restrictions on the original model, $\delta \in \Delta$ and $\beta = \beta^*$, echo into the reduced model.

The reduced model is a classical linear regression. We assume a Gaussian distribution with parameters ρ, Γ and Σ.

Application to Sensitivity Analysis

Original model

$$Y = D\beta + Z\delta + \varepsilon_Y,$$
$$D = Z\Gamma + \varepsilon_D.$$

Reduced model

$$Y = Z\rho + \varepsilon_Y,$$
$$D = Z\Gamma + \varepsilon_D.$$

The original and reduced model are linked via $\rho = \delta + \Gamma\beta$. Hence, the restrictions on the original model, $\delta \in \Delta$ and $\beta = \beta^*$, echo into the reduced model.

The reduced model is a classical linear regression. We assume a Gaussian distribution with parameters ρ, Γ and Σ.

Application to Sensitivity Analysis

Parameters spaces for testing H_0: $\beta = \beta^*$:

$$\Theta := \{(\Gamma, \rho) \in \mathbb{R}^{k \times p} \times \mathbb{R}^k\},$$
$$\Theta_1 := \{(\Gamma, \rho) \in \mathbb{R}^{k \times p} \times \mathbb{R}^k \mid \exists\, \delta \in \Delta\, \exists\, \beta \in \mathbb{R}^p \colon \rho = \delta + \Gamma\beta\},$$
$$\Theta_0 := \{(\Gamma, \rho) \in \mathbb{R}^{k \times p} \times \mathbb{R}^k \mid \exists\, \delta \in \Delta \colon \rho = \delta + \Gamma\beta^*\}.$$

In Gaussian case, we can replace Σ with a consistent estimate $\widehat{\Sigma}$.

- Θ_0 vs. Θ_1: Assumption that sensitivity model is correctly specified; always non-empty sensitivity region; difficult limit distribution
- Θ_0 vs. Θ: Test for $\beta = \beta^*$ and correctness of sensitivity model; empty sensitivity region possible; easier limit distribution

Application to Sensitivity Analysis

Parameters spaces for testing H_0: $\beta = \beta^*$:

$$\begin{aligned}
\Theta &:= \{(\Gamma, \rho) \in \mathbb{R}^{k \times p} \times \mathbb{R}^k\}, \\
\Theta_1 &:= \{(\Gamma, \rho) \in \mathbb{R}^{k \times p} \times \mathbb{R}^k \mid \exists\, \delta \in \Delta\, \exists\, \beta \in \mathbb{R}^p \colon \rho = \delta + \Gamma\beta\}, \\
\Theta_0 &:= \{(\Gamma, \rho) \in \mathbb{R}^{k \times p} \times \mathbb{R}^k \mid \exists\, \delta \in \Delta \colon \rho = \delta + \Gamma\beta^*\}.
\end{aligned}$$

In Gaussian case, we can replace Σ with a consistent estimate $\widehat{\Sigma}$.

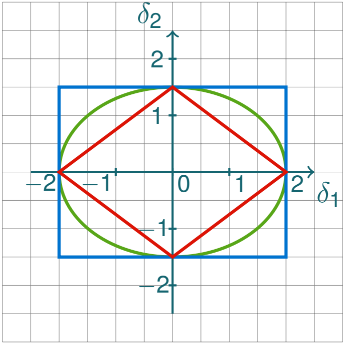
- Θ_0 vs. Θ_1: Assumption that sensitivity model is correctly specified; always non-empty sensitivity region; difficult limit distribution
- Θ_0 vs. Θ: Test for $\beta = \beta^*$ and correctness of sensitivity model; empty sensitivity region possible; easier limit distribution

Sensitivity Sets

Goal: easy specification for user and consideration of limit distribution for special cases

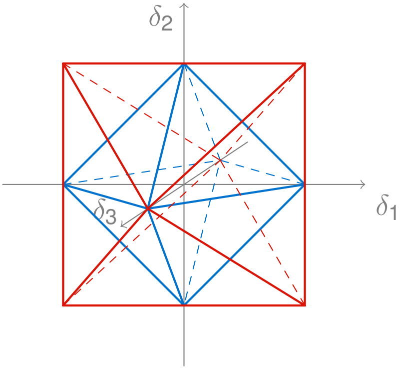
Range definition: For every instrument Z_j the user stipulates a range $[\delta_l^{(j)}, \delta_u^{(j)}]$ and chooses "interpolation" between instruments

- Ellipsoid (L^2-ball like): limit distribution $0.5\chi_1^2 + 0.5\chi_0^2$
- Hypercube (L^∞-ball like): finite number of least favourable nulls
- Cross polytope (L^1-ball like): finite number of least favourable nulls

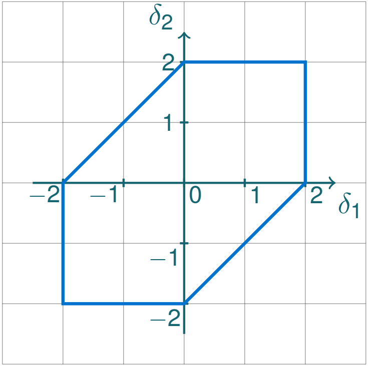


Sensitivity Sets - Combinations and Categorical IVs

Combination of hypercube and cross-polytope interpolation



Sensitivity set for a 3-level categorical instrument



Outlook

Short-term:

- Finish work on more complex sensitivity sets
- Implementation and empirical evaluation

Mid-term:

- R-package
- Moderate generalisation: simultaneous equations, spline IV
- Connection between constrained inference and post-selection inference literature

Long-term:

- Further theoretical development of adaptive constrained inference, cf. Al Mohamad et al. (2020)
- Major generalisation: semiparametric IV, kernel IV

References I

Al Mohamad, Diaa, Erik W Van Zwet, Eric Cator, and Jelle J Goeman (2020). "Adaptive critical value for constrained likelihood ratio testing". In: *Biometrika* 107.3, pp. 677–688.

Baiocchi, Michael, Jing Cheng, and Dylan S. Small (2014). "Instrumental variable methods for causal inference". In: *Statistics in Medicine* 33.13, pp. 2297–2340.

Card, David (1995). "The Wage Curve: A Review". In: *Journal of Economic Literature* 33.2, pp. 785–799.

Fukuda, Komei (2020). *Polyhedral computation*. Tech. rep. Department of Mathematics, Institute of Theoretical Computer Science ETH Zurich.

Silvapulle, Mervyn J. and Pranab K. Sen (2005). *Constrained statistical inference. Inequality, order, and shape restrictions.* Hoboken, NJ: John Wiley & Sons, pp. xvii + 532.

Wang, Xuran, Yang Jiang, Nancy R. Zhang, and Dylan S. Small (2018). "Sensitivity analysis and power for instrumental variable studies". In: *Biometrics* 74.4, pp. 1150–1160.

References II

Wasserman, Larry, Aaditya Ramdas, and Sivaraman Balakrishnan (2020). "Universal inference". In: *Proceedings of the National Academy of Sciences* 117.29, pp. 16880–16890.

Wooldridge, Jefrey M. (2010). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press.

Zhao, Qingyuan, Dylan S. Small, and Bhaswar B. Bhattacharya (2018). *Sensitivity analysis for inverse probability weighting estimators via the percentile bootstrap*. arXiv: 1711.11286.