

Mis-Classified, Binary, Endogenous Regressors: Identification and Inference

Francis J. DiTraglia¹ Camilo García-Jimeno^{2,3}

¹University of Pennsylvania

²Emory University

³NBER

November 14th, 2018

What is the effect of T^* ?

$$y = c(\mathbf{x}) + \beta(\mathbf{x})T^* + \varepsilon$$

- ▶ y – Outcome of interest
- ▶ T^* – Unobserved, endogenous binary regressor
- ▶ T – Observed, mis-measured binary surrogate for T^*
- ▶ \mathbf{x} – Exogenous covariates
- ▶ z – Discrete (typically binary) instrumental variable

Using a discrete IV to learn about $\beta(\mathbf{x})$

$$y = c(\mathbf{x}) + \beta(\mathbf{x})T^* + \varepsilon$$

Contributions of This Paper

1. Show that only existing point identification result for mis-classified, endogenous T^* is incorrect.
2. Sharp identified set for β under standard assumptions.
3. Point identification of β under slightly stronger assumptions.
4. Describe problem of weak identification in mis-classification models, develop identification-robust inference for β .

Example: Smoking and Birthweight (SNAP Trial)

Coleman et al. (N Engl J Med, 2012)

RCT with pregnant smokers in England: half given nicotine patches, the rest given placebo patches. Some given nicotine fail to quit; some given placebo quit.

- ▶ y – Birthweight
- ▶ T^* – True smoking behavior
- ▶ T – Self-reported smoking behavior
- ▶ x – Mother characteristics
- ▶ z – Indicator of nicotine patch

Baseline Assumptions I – Model & Instrument

Additively Separable Model

$$y = c(\mathbf{x}) + \beta(\mathbf{x})T^* + \varepsilon, \quad \mathbb{E}[\varepsilon] = 0$$

Valid & Relevant Instrument: $z \in \{0, 1\}$

- ▶ $\mathbb{P}(T^* = 1|\mathbf{x}, z = 1) \neq \mathbb{P}(T^* = 1|\mathbf{x}, z = 0)$
- ▶ $\mathbb{E}[\varepsilon|\mathbf{x}, z] = 0$
- ▶ $0 < \mathbb{P}(z = 1|\mathbf{x}) < 1$

Baseline Assumptions II – Measurement Error

Notation

► $\alpha_0(\mathbf{x}, z) \equiv \mathbb{P}(T = 1 | T^* = 0, \mathbf{x}, z)$

► $\alpha_1(\mathbf{x}, z) \equiv \mathbb{P}(T = 0 | T^* = 1, \mathbf{x}, z)$

Mis-classification unaffected by z

$$\alpha_0(\mathbf{x}, z) = \alpha_0(\mathbf{x}), \quad \alpha_1(\mathbf{x}, z) = \alpha_1(\mathbf{x})$$

Extent of Mis-classification

$$\alpha_0(\mathbf{x}) + \alpha_1(\mathbf{x}) < 1 \quad (T \text{ is positively correlated with } T^*)$$

Non-differential Mis-classification

$$\mathbb{E}[\varepsilon | \mathbf{x}, z, T^*, T] = \mathbb{E}[\varepsilon | \mathbf{x}, z, T^*]$$

Only Existing Result for Endogenous T^* is Incorrect

Mahajan (2006; Ecta) A2

$$\mathbb{E}[\varepsilon|\mathbf{x}, z, T^*, T] = \mathbb{E}[\varepsilon|\mathbf{x}, T^*] + \text{"Baseline"} \Rightarrow \beta(\mathbf{x}) \text{ identified.}$$

We Show:

Mahajan's assumptions imply that the instrument z is uncorrelated with T^* unless T^* is in fact *exogenous*.

“Weak” Bounds

First-Stage

$$p_k(\mathbf{x}) \equiv \mathbb{P}(T = 1 | \mathbf{x}, z = k)$$

IV Estimand

$$\frac{\mathbb{E}[y | \mathbf{x}, z = 1] - \mathbb{E}[y | \mathbf{x}, z = 0]}{p_1(\mathbf{x}) - p_0(\mathbf{x})} = \frac{\beta(\mathbf{x})}{1 - \alpha_0(\mathbf{x}) - \alpha_1(\mathbf{x})}$$

Bounds for (α_0, α_1)

$$\alpha_0(\mathbf{x}) \leq \min_k \{p_k(\mathbf{x})\}, \quad \alpha_1(\mathbf{x}) \leq \min_k \{1 - p_k(\mathbf{x})\} \quad \text{▶ more}$$

Bounds for β

$\beta(\mathbf{x})$ is between IV and Reduced form; same sign as IV. ▶ more

Restrictions from Non-differential Mis-classification?

(Suppress \mathbf{x} for simplicity)

Notation

- ▶ $r_{tk} \equiv \mathbb{P}(T^* = 1 | T = t, z = k)$
- ▶ z_k is shorthand for $z = k$

Iterated Expectations over T^*

$$\mathbb{E}(y | T = 0, z_k) = (1 - r_{0k})\mathbb{E}(y | T^* = 0, T = 0, z_k) + r_{0k}\mathbb{E}(y | T^* = 1, T = 0, z_k)$$

$$\mathbb{E}(y | T = 1, z_k) = (1 - r_{1k})\mathbb{E}(y | T^* = 0, T = 1, z_k) + r_{1k}\mathbb{E}(y | T^* = 1, T = 1, z_k)$$

Restrictions from Non-differential Mis-classification?

(Suppress \mathbf{x} for simplicity)

Notation

- ▶ $r_{tk} \equiv \mathbb{P}(T^* = 1 | T = t, z = k)$
- ▶ z_k is shorthand for $z = k$

Adding Non-differential Assumption

$$\mathbb{E}(y | T = 0, z_k) = (1 - r_{0k})\mathbb{E}(y | T^* = 0, z_k) + r_{0k}\mathbb{E}(y | T^* = 1, z_k)$$

$$\mathbb{E}(y | T = 1, z_k) = (1 - r_{1k})\mathbb{E}(y | T^* = 0, z_k) + r_{1k}\mathbb{E}(y | T^* = 1, z_k)$$

2 equations in 2 unknowns \Rightarrow solve for $\mathbb{E}(y T^* = t^*, z = k)$ given (r_{0k}, r_{1k}) .

Restrictions from Non-differential Mis-classification?

Mixture Representation

$$F_{tk} = (1 - r_{tk})F_{tk}^0 + r_{tk}F_{tk}^1$$

$$F_{tk} \equiv y | (T = t, z = k)$$

$$F_{tk}^{t^*} \equiv y | (T^* = t^*, T = t, z = k)$$

Restrictions

- ▶ $\mathbb{E}(y | T^*, T, z) = \mathbb{E}(y | T^*, z)$ observable given (α_0, α_1)
- ▶ r_{tk} observable given (α_0, α_1)

Question

Given (α_0, α_1) can we always find (F_{tk}^0, F_{tk}^1) to satisfy the mixture model?

Restrictions from Non-differential Mis-classification?

Equivalent Problem

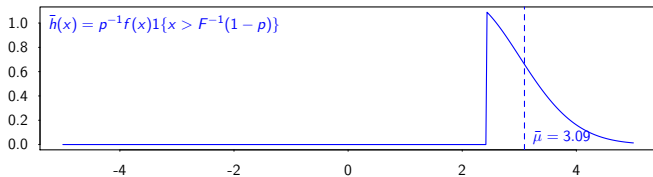
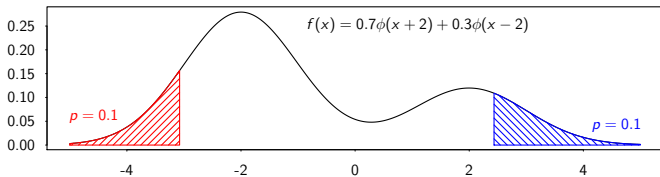
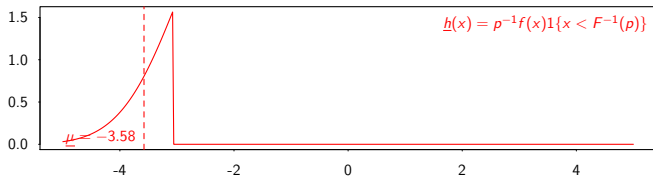
Given a specified CDF F , for what values of p and μ do there exist valid CDFs (G, H) with $F = (1 - p)G + pH$ and $\mu = \text{mean}(H)$?

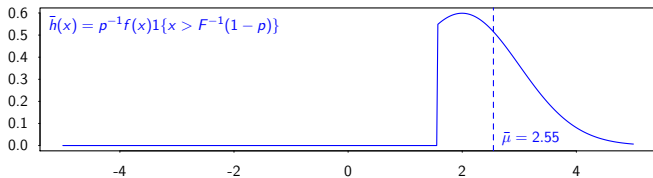
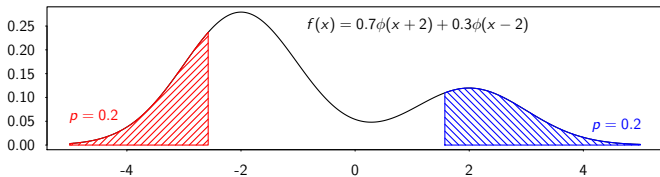
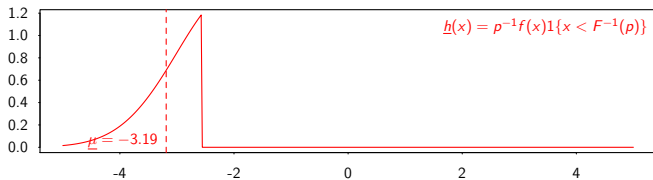
Necessary and Sufficient Condition if F is Continuous

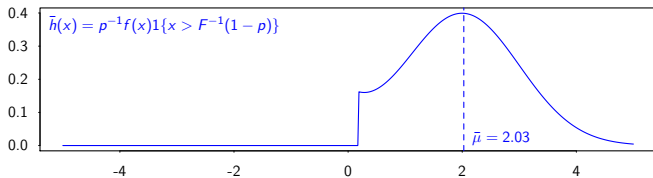
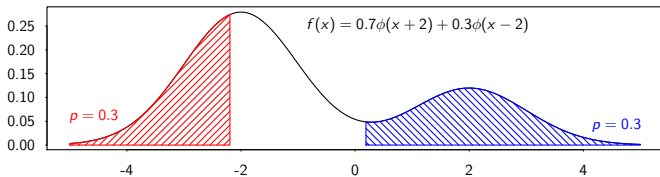
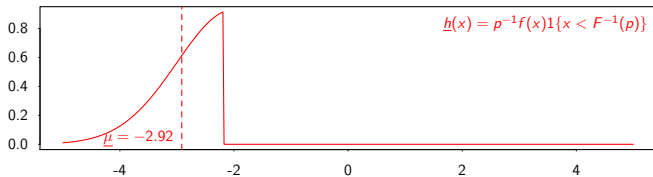
$$\underline{\mu}(F, p) \leq \mu \leq \bar{\mu}(F, p)$$

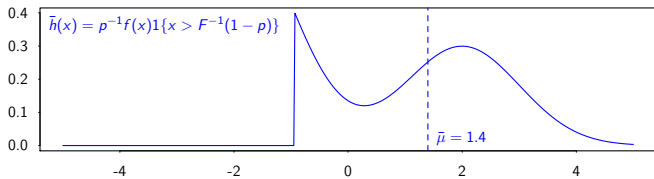
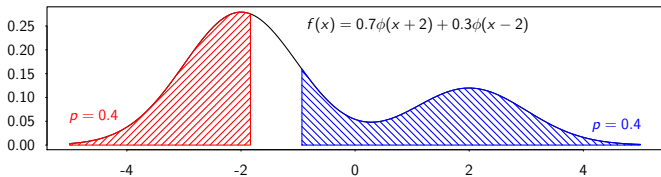
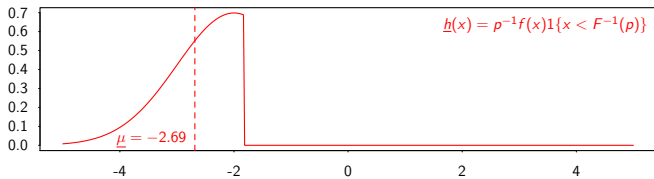
$$\underline{\mu}(F, p) \equiv \int_{-\infty}^{\infty} x \left[p^{-1} f(x) \mathbf{1}\{x < F^{-1}(p)\} \right] dx = \int_{-\infty}^{\infty} x \underline{h}(x) dx$$

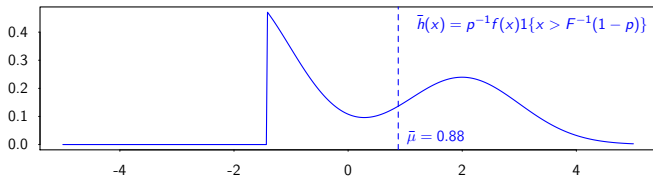
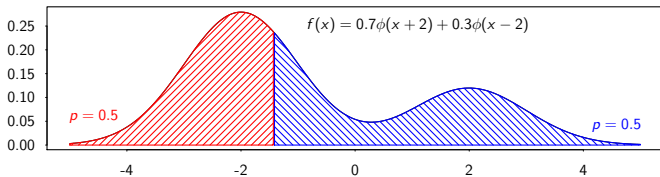
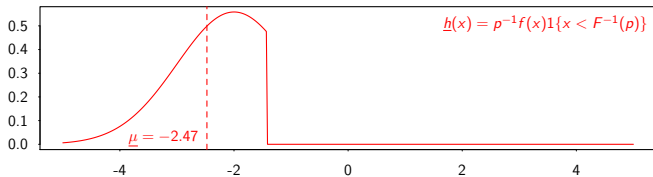
$$\bar{\mu}(F, p) \equiv \int_{-\infty}^{\infty} x \left[p^{-1} f(x) \mathbf{1}\{x > F^{-1}(1 - p)\} \right] dx = \int_{-\infty}^{\infty} x \bar{h}(x) dx$$

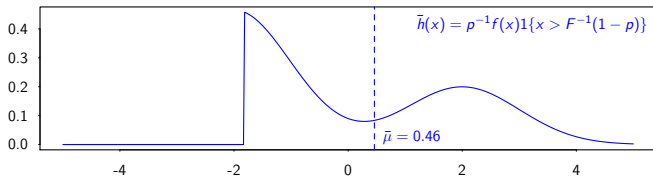
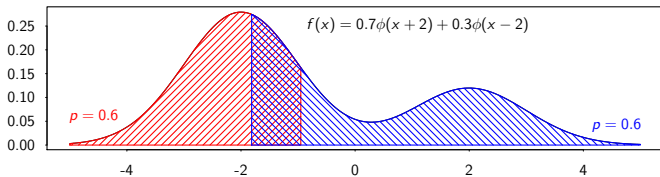
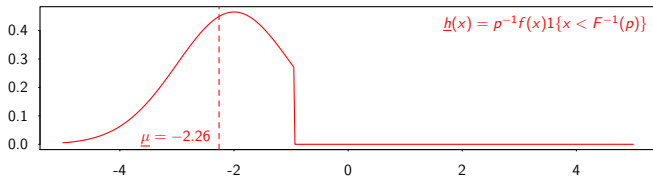


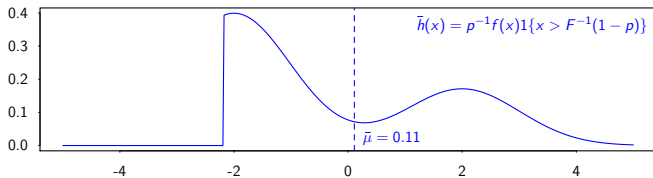
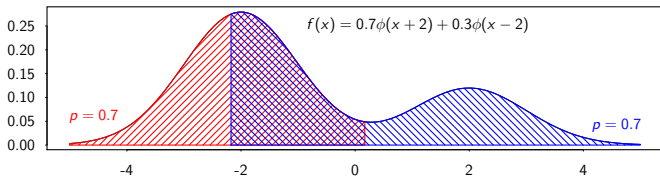
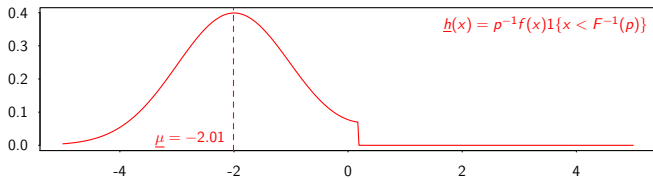


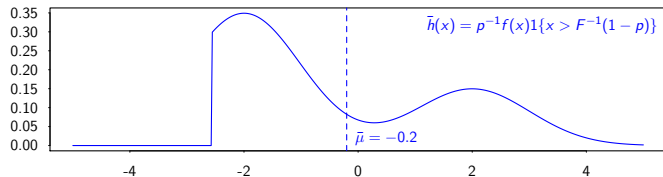
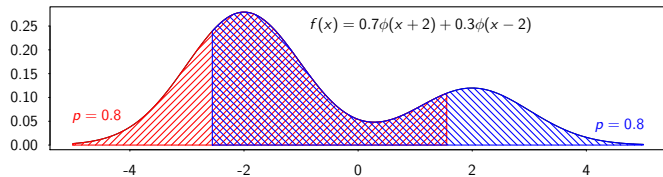
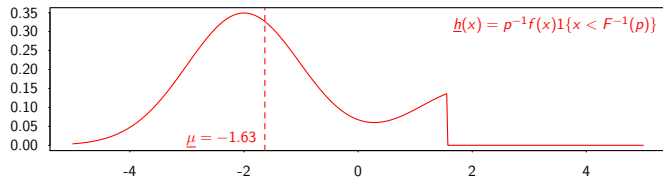


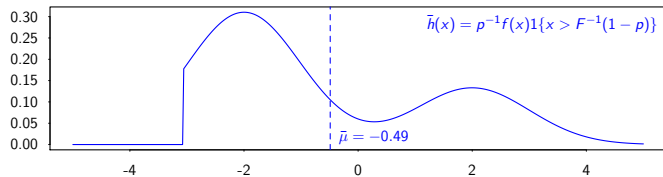
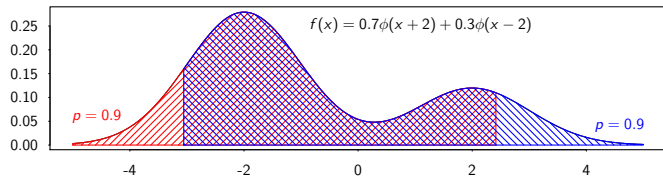
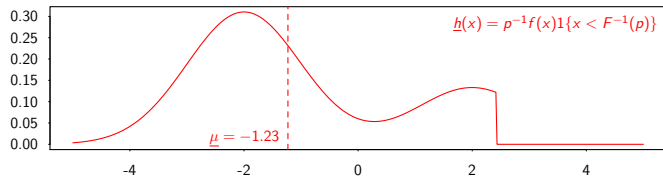


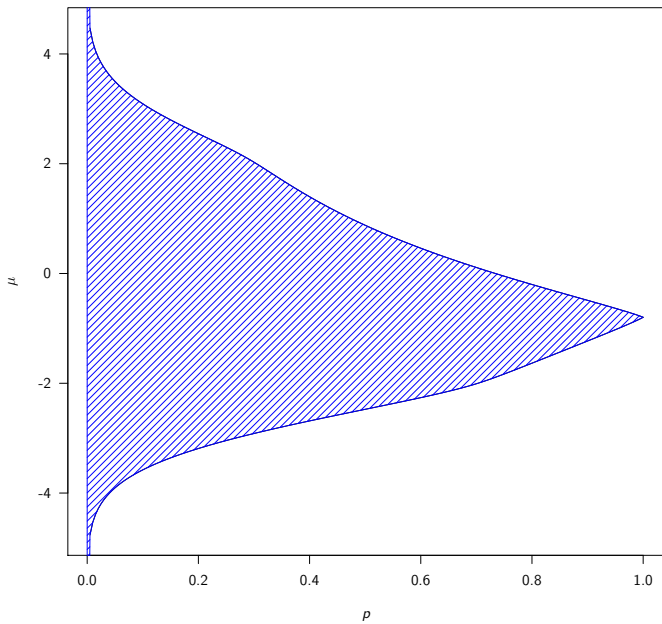












Sharp Identified Set under Baseline Assumptions

Theorem

- (i) If $\mathbb{E}[y|\mathbf{x}, T = 0, z = k] \neq \mathbb{E}[y|\mathbf{x}, T = 1, z = k]$ for some k , non-differential assump. strictly improves upon weak bounds.
- (ii) Under the baseline assumptions, β is not point identified, regardless of how many (discrete) values z takes on.

Corollary

Bounds for α_0, α_1 , and β remain valid in a LATE model. They may not be sharp, however, sharp, since they do not incorporate the testable implications of the LATE assumptions.

Point Identification: 1st Ingredient

Reparameterization

$$\theta_1(\mathbf{x}) = \beta(\mathbf{x}) / [1 - \alpha_0(\mathbf{x}) - \alpha_1(\mathbf{x})]$$

$$\theta_2(\mathbf{x}) = [\theta_1(\mathbf{x})]^2 [1 + \alpha_0(\mathbf{x}) - \alpha_1(\mathbf{x})]$$

$$\theta_3(\mathbf{x}) = [\theta_1(\mathbf{x})]^3 \left[\{1 - \alpha_0(\mathbf{x}) - \alpha_1(\mathbf{x})\}^2 + 6\alpha_0(\mathbf{x}) \{1 - \alpha_1(\mathbf{x})\} \right]$$

Lemma

Baseline Assumptions $\implies \text{Cov}(y, z|\mathbf{x}) = \theta_1(\mathbf{x})\text{Cov}(z, T|\mathbf{x})$.

Point Identification: 2nd Ingredient

Assumption (II)

$$\mathbb{E}[\varepsilon^2|\mathbf{x}, z] = \mathbb{E}[\varepsilon^2|\mathbf{x}]$$

Lemma

$$(\text{Baseline}) + (\text{II}) \implies$$

$$\text{Cov}(y^2, z|\mathbf{x}) = 2\text{Cov}(yT, z|\mathbf{x})\theta_1(\mathbf{x}) - \text{Cov}(T, z|\mathbf{x})\theta_2(\mathbf{x})$$

Corollary

$$(\text{Baseline}) + (\text{II}) + [\beta(\mathbf{x}) \neq 0] \implies [\alpha_1(\mathbf{x}) - \alpha_0(\mathbf{x})] \text{ is identified.}$$

Point Identification: 3rd Ingredient

Assumption (III)

$$(i) \mathbb{E}[\varepsilon^2 | \mathbf{x}, z, T^*, T] = \mathbb{E}[\varepsilon^2 | \mathbf{x}, z, T^*]$$

$$(ii) \mathbb{E}[\varepsilon^3 | \mathbf{x}, z] = \mathbb{E}[\varepsilon^3 | \mathbf{x}]$$

Lemma

(Baseline) + (II) + (III) \implies

$$\text{Cov}(y^3, z | \mathbf{x}) = 3\text{Cov}(y^2 T, z | \mathbf{x})\theta_1(\mathbf{x}) - 3\text{Cov}(yT, z | \mathbf{x})\theta_2(\mathbf{x}) + \text{Cov}(T, z | \mathbf{x})\theta_3(\mathbf{x})$$

Point Identification Result

Theorem

(Baseline) + (II) + (III) $\implies \beta(\mathbf{x})$ is point identified. If $\beta(\mathbf{x}) \neq 0$, then $\alpha_0(\mathbf{x})$ and $\alpha_1(\mathbf{x})$ are likewise point identified.

Sufficient for (II) and (III)

- (a) T is conditionally independent of (ε, z) given (T^*, \mathbf{x})
- (b) z is conditionally independent of ε given \mathbf{x}

Inference for a Mis-classified Regressor

Challenges

- ▶ Weak Identification: β small \Rightarrow moment equalities uninformative about (α_0, α_1) [▶ more](#)
- ▶ (α_0, α_1) could be on the boundary of the parameter space
- ▶ Also true of existing estimators that assume T^* exogenous

Our Approach

- ▶ Sharp identified set yields *inequality* moment restrictions that remain informative even if $\beta \approx 0$. [▶ more](#)
- ▶ Identification-robust inference with equality and inequality MCs.

Inference with Moment Equalities and Inequalities

Moment Conditions

$$\mathbb{E}[m_j(\mathbf{w}_i, \vartheta_0)] \geq 0, \quad j = 1, \dots, J$$

$$\mathbb{E}[m_j(\mathbf{w}_i, \vartheta_0)] = 0, \quad j = J+1, \dots, J+K$$

Test Statistic

$$T_n(\vartheta) = \sum_{j=1}^J \left[\frac{\sqrt{n} \bar{m}_{n,j}(\vartheta)}{\hat{\sigma}_{n,j}(\vartheta)} \right]_-^2 + \sum_{j=J+1}^{J+K} \left[\frac{\sqrt{n} \bar{m}_{n,j}(\vartheta)}{\hat{\sigma}_{n,j}(\vartheta)} \right]^2$$

Critical Value

- ▶ $\sqrt{n} \bar{m}_n(\vartheta_0) \rightarrow_d$ normal limit with covariance matrix $\Sigma(\vartheta_0)$
- ▶ Use this to bootstrap the limit dist. of $T_n(\vartheta)$ under $H_0: \vartheta = \vartheta_0$

Generalized Moment Selection

Andrews & Soares (2010)

- ▶ Inequalities that don't bind reduce power of test, so eliminate those that are “far from binding” before calculating critical value:

$$\text{Drop inequality } j \text{ if } \frac{\sqrt{n} \bar{m}_{n,j}(\vartheta_0)}{\hat{\sigma}_{n,j}(\vartheta_0)} > \sqrt{\log n}$$

- ▶ Uniformly valid test of $H_0: \vartheta = \vartheta_0$ even if ϑ_0 is not point identified.
- ▶ Not asymptotically conservative.

Problem

Joint test for the whole parameter vector but we're only interested in β .
Projection is conservative and computationally intensive.

Our Solution: Bonferroni-Based Inference

Special Structure

- ▶ β only enters MCs through $\theta_1 = \beta/(1 - \alpha_0 - \alpha_1)$
- ▶ Strong instrument \Rightarrow inference for θ_1 is standard.
- ▶ Nuisance pars γ strongly identified under null for (α_0, α_1)

Our Solution: Bonferroni-Based Inference

Special Structure

- ▶ β only enters MCs through $\theta_1 = \beta/(1 - \alpha_0 - \alpha_1)$
- ▶ Strong instrument \Rightarrow inference for θ_1 is standard.
- ▶ Nuisance pars γ strongly identified under null for (α_0, α_1)

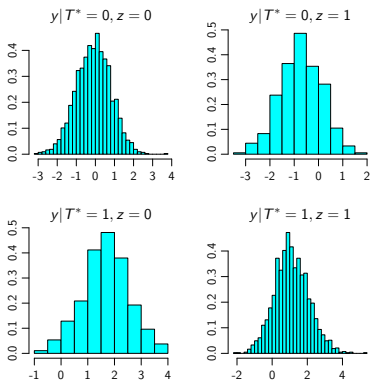
Procedure

1. Concentrate out $(\theta_1, \gamma) \Rightarrow$ joint GMS test for (α_0, α_1)
2. Invert test $\Rightarrow (1 - \delta_1) \times 100\%$ confidence set for (α_0, α_1)
3. Project \Rightarrow CI for $(1 - \alpha_0 - \alpha_1)$
4. Construct standard $(1 - \delta_2) \times 100\%$ IV CI for θ_1
5. Bonferroni $\Rightarrow (1 - \delta_1 - \delta_2) \times 100\%$ CI for β

Example

(sim data: $\beta = 1, \alpha_0 = 0.1, \alpha_1 = 0.2, n = 5000$)

Results if T^* were observed

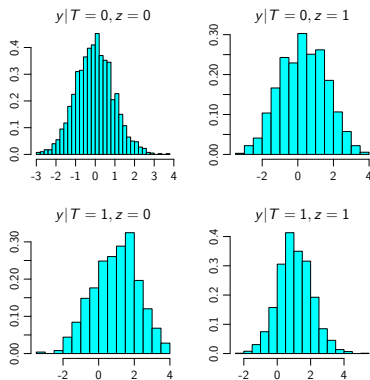


$$\hat{\beta}_{IV} = 0.96, \quad 95\% \text{ CI} = (0.88, 1.04)$$

Example

(sim data: $\beta = 1, \alpha_0 = 0.1, \alpha_1 = 0.2, n = 5000$)

Results using T instead of T^*

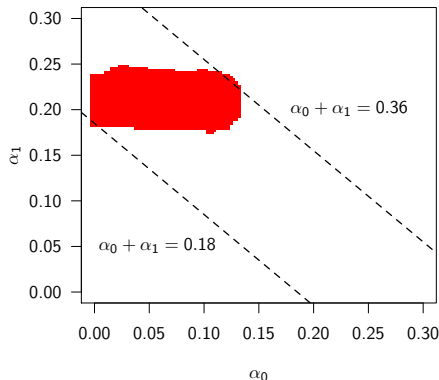


$$\hat{\beta}_{IV} = 1.34, \quad 95\% \text{ CI} = (1.22, 1.45)$$

Example

(sim data: $\beta = 1, \alpha_0 = 0.1, \alpha_1 = 0.2, n = 5000$)

97.5% GMS Confidence Region for (α_0, α_1) Our Procedure

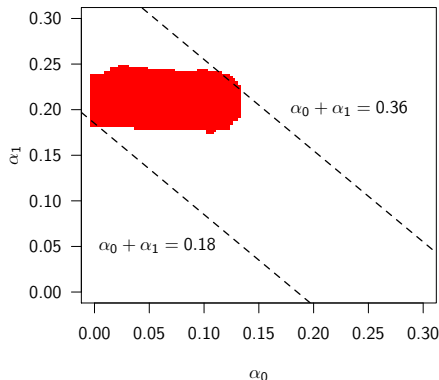


1. 97.5% GMS CI for (α_0, α_1) at left

Example

(sim data: $\beta = 1, \alpha_0 = 0.1, \alpha_1 = 0.2, n = 5000$)

97.5% GMS Confidence Region for (α_0, α_1) Our Procedure

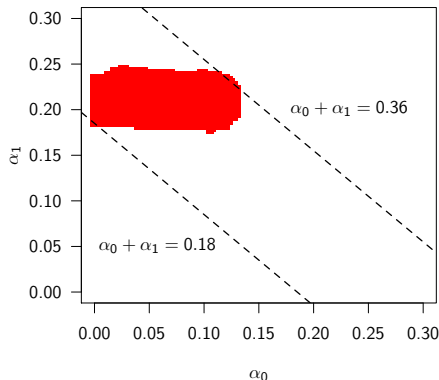


1. 97.5% GMS CI for (α_0, α_1) at left
2. $\geq 97.5\%$ Projection CI for $(1 - \alpha_0 - \alpha_1)$
 $(1 - 0.36, 1 - 0.18) = (0.64, 0.82)$

Example

(sim data: $\beta = 1, \alpha_0 = 0.1, \alpha_1 = 0.2, n = 5000$)

97.5% GMS Confidence Region for (α_0, α_1) Our Procedure

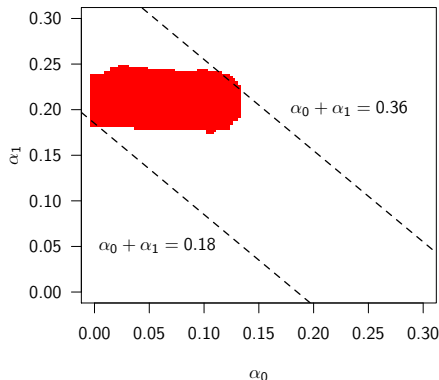


1. 97.5% GMS CI for (α_0, α_1) at left
2. $\geq 97.5\%$ Projection CI for $(1 - \alpha_0 - \alpha_1)$
 $(1 - 0.36, 1 - 0.18) = (0.64, 0.82)$
3. 97.5% CI for $\theta_1 = (1.20, 1.47)$

Example

(sim data: $\beta = 1, \alpha_0 = 0.1, \alpha_1 = 0.2, n = 5000$)

97.5% GMS Confidence Region for (α_0, α_1) Our Procedure

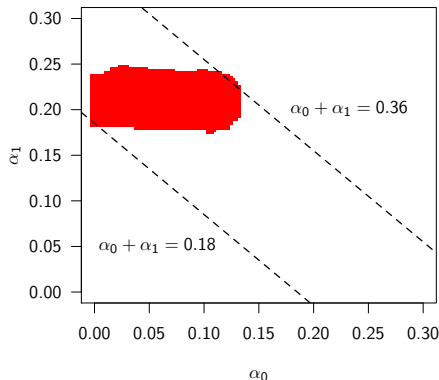


1. 97.5% GMS CI for (α_0, α_1) at left
2. $\geq 97.5\%$ Projection CI for $(1 - \alpha_0 - \alpha_1)$
 $(1 - 0.36, 1 - 0.18) = (0.64, 0.82)$
3. 97.5% CI for $\theta_1 = (1.20, 1.47)$
4. $\geq 95\%$ CI for β :
 $(0.64 \times 1.20, 0.82 \times 1.47) = (0.77, 1.21)$

Example

(sim data: $\beta = 1, \alpha_0 = 0.1, \alpha_1 = 0.2, n = 5000$)

97.5% GMS Confidence Region for (α_0, α_1) Our Procedure



1. 97.5% GMS CI for (α_0, α_1) at left
2. $\geq 97.5\%$ Projection CI for $(1 - \alpha_0 - \alpha_1)$
 $(1 - 0.36, 1 - 0.18) = (0.64, 0.82)$
3. 97.5% CI for $\theta_1 = (1.20, 1.47)$
4. $\geq 95\%$ CI for β :
 $(0.64 \times 1.20, 0.82 \times 1.47) = (0.77, 1.21)$

Comparisons

- ▶ $(0.88, 1.04)$ 95% CI for IV using T^*
- ▶ $(1.22, 1.45)$ 95% CI for IV using T

Conclusion

This Paper

- ▶ Partial and point identification results for effect of binary, endogenous regressor using a valid instrument.
- ▶ Identification-robust inference in models with mis-classification

Related Work

- ▶ Relaxing Instrument Validity: “A Framework for Eliticing, Incorporating, and Disciplining Identification Beliefs in Linear Models” (with Camilo Garcia-Jimeno)
- ▶ Relaxing Non-differential Measurement Error: “Estimating the Returns to Lying” (with Arthur Lewbel)

Simple Bounds for Mis-classification from First-stage

Unobserved	Observed
$p_k^*(\mathbf{x}) \equiv \mathbb{P}(T^* = 1 \mathbf{x}, z = k)$	$p_k(\mathbf{x}) \equiv \mathbb{P}(T = 1 \mathbf{x}, z = k)$

Relationship

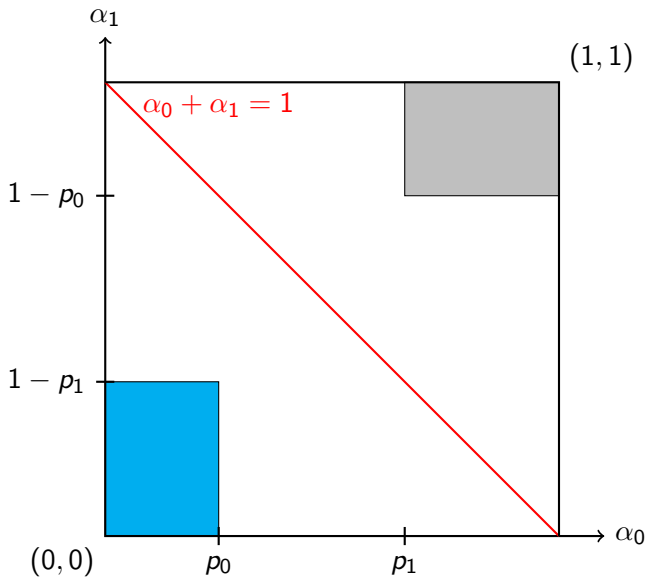
$$p_k^*(\mathbf{x}) = \frac{p_k(\mathbf{x}) - \alpha_0(\mathbf{x})}{1 - \alpha_0(\mathbf{x}) - \alpha_1(\mathbf{x})}, \quad k = 0, 1$$

z does not affect (α_0, α_1) ; denominator $\neq 0$

Bounds for Mis-classification

$$\alpha_0(\mathbf{x}) \leq p_k(\mathbf{x}) \leq 1 - \alpha_1(\mathbf{x}), \quad k = 0, 1$$

$$\alpha_0(\mathbf{x}) + \alpha_1(\mathbf{x}) < 1$$



What does IV estimate under mis-classification?

Unobserved

$$\beta(\mathbf{x}) = \frac{\mathbb{E}[y|\mathbf{x}, z = 1] - \mathbb{E}[y|\mathbf{x}, z = 0]}{p_1^*(\mathbf{x}) - p_0^*(\mathbf{x})}$$

Wald (Observed)

$$\frac{\mathbb{E}[y|\mathbf{x}, z = 1] - \mathbb{E}[y|\mathbf{x}, z = 0]}{p_1(\mathbf{x}) - p_0(\mathbf{x})} = \beta(\mathbf{x}) \left[\frac{p_1^*(\mathbf{x}) - p_0^*(\mathbf{x})}{p_1(\mathbf{x}) - p_0(\mathbf{x})} \right] = \frac{\beta(\mathbf{x})}{1 - \alpha_0(\mathbf{x}) - \alpha_1(\mathbf{x})}$$

$$p_1^*(\mathbf{x}) - p_0^*(\mathbf{x}) = \frac{p_1(\mathbf{x}) - \alpha_0(\mathbf{x})}{1 - \alpha_0 - \alpha_1(\mathbf{x})} - \frac{p_0(\mathbf{x}) - \alpha_0(\mathbf{x})}{1 - \alpha_0 - \alpha_1(\mathbf{x})} = \frac{p_1(\mathbf{x}) - p_0(\mathbf{x})}{1 - \alpha_0(\mathbf{x}) - \alpha_1(\mathbf{x})}$$

Partial Identification Bounds for $\beta(\mathbf{x})$

$$\beta(\mathbf{x}) = [1 - \alpha_0(\mathbf{x}) - \alpha_1(\mathbf{x})] \left[\frac{\mathbb{E}[y|\mathbf{x}, z = 1] - \mathbb{E}[y|\mathbf{x}, z = 0]}{p_1(\mathbf{x}) - p_0(\mathbf{x})} \right]$$

$$0 \leq \alpha_0 \leq \min_k \{p_k(\mathbf{x})\}, \quad 0 \leq \alpha_1 \leq \min_k \{1 - p_k(\mathbf{x})\}$$

No Mis-classification

$$\alpha_0(\mathbf{x}) = \alpha_1(\mathbf{x}) = 0 \implies \beta(\mathbf{x}) = \text{Wald}$$

Maximum Mis-classification

$$\alpha_0(\mathbf{x}) = p_{\min}(\mathbf{x}), \quad \alpha_1(\mathbf{x}) = 1 - p_{\max}(\mathbf{x})$$

$$\implies 1 - \alpha_0(\mathbf{x}) - \alpha_1(\mathbf{x}) = p_{\max}(\mathbf{x}) - p_{\min}(\mathbf{x}) = |p_1(\mathbf{x}) - p_0(\mathbf{x})|$$

$$\implies \beta(\mathbf{x}) = \text{sign}\{p_1(\mathbf{x}) - p_0(\mathbf{x})\} \times (\text{Reduced Form})$$

Just-Identified System of Moment Equalities

Suppress dependence on $\mathbf{x} \dots$

$$\mathbb{E} \left[\{ \boldsymbol{\Psi}(\boldsymbol{\theta}) \mathbf{w}_i - \boldsymbol{\kappa} \} \otimes \begin{pmatrix} 1 \\ z \end{pmatrix} \right] = \mathbf{0}$$

$$\boldsymbol{\Psi}(\boldsymbol{\theta}) \equiv \begin{bmatrix} -\theta_1 & 1 & 0 & 0 & 0 & 0 \\ \theta_2 & 0 & -2\theta_1 & 1 & 0 & 0 \\ -\theta_3 & 0 & 3\theta_2 & 0 & -3\theta_1 & 1 \end{bmatrix}$$

$$\mathbf{w}_i = (T_i, y_i, y_i T_i, y_i^2, y_i^2 T_i, y_i^3)' \quad \theta_1 = \beta / (1 - \alpha_0 - \alpha_1)$$

$$\boldsymbol{\kappa} = (\kappa_1, \kappa_2, \kappa_3)' \quad \theta_2 = \theta_1^2 (1 + \alpha_0 - \alpha_1)$$

$$\theta_3 = \theta_1^3 [(1 - \alpha_0 - \alpha_1)^2 + 6\alpha_0(1 - \alpha_1)]$$

Moment Inequalities I – First-stage Probabilities

$\alpha_0 \leq p_k \leq 1 - \alpha_1$ becomes $\mathbb{E}[m(\mathbf{w}_i, \boldsymbol{\vartheta})] \geq \mathbf{0}$ for all k where

$$m(\mathbf{w}_i, \boldsymbol{\vartheta}) \equiv \begin{bmatrix} \mathbf{1}(z_i = k)(T - \alpha_0) \\ \mathbf{1}(z_i = k)(1 - T_i - \alpha_1) \end{bmatrix}$$

Moment Inequalities II – Non-differential Assumption

For all k , we have $\mathbb{E}[m(\mathbf{w}_i, \vartheta, \mathbf{q}_k)] \geq 0$ where

$$m(\mathbf{w}_i, \vartheta, \mathbf{q}_k) \equiv \begin{bmatrix} y_i \mathbf{1}(z_i = k) \left\{ (T_i - \alpha_0) - \mathbf{1}(y_i \leq \underline{q}_{0k})(1 - T_i) \left(\frac{1 - \alpha_0 - \alpha_1}{\alpha_1} \right) \right\} \\ -y_i \mathbf{1}(z_i = k) \left\{ (T_i - \alpha_0) - \mathbf{1}(y_i > \bar{q}_{0k})(1 - T_i) \left(\frac{1 - \alpha_0 - \alpha_1}{\alpha_1} \right) \right\} \\ y_i \mathbf{1}(z_i = k) \left\{ (T_i - \alpha_0) - \mathbf{1}(y_i \leq \underline{q}_{1k}) T_i \left(\frac{1 - \alpha_0 - \alpha_1}{1 - \alpha_1} \right) \right\} \\ -y_i \mathbf{1}(z_i = k) \left\{ (T_i - \alpha_0) - \mathbf{1}(y_i > \bar{q}_{1k}) T_i \left(\frac{1 - \alpha_0 - \alpha_1}{1 - \alpha_1} \right) \right\} \end{bmatrix}$$

and $\mathbf{q}_k \equiv (\underline{q}_{0k}, \bar{q}_{0k}, \underline{q}_{1k}, \bar{q}_{1k})'$ defined by $\mathbb{E}[h(\mathbf{w}_i, \vartheta, \mathbf{q}_k)] = 0$ with

$$h(\mathbf{w}_i, \vartheta, \mathbf{q}_k) = \begin{bmatrix} \mathbf{1}(y_i \leq \underline{q}_{0k}) \mathbf{1}(z_i = k)(1 - T_i) - \left(\frac{\alpha_1}{1 - \alpha_0 - \alpha_1} \right) \mathbf{1}(z_i = k)(T_i - \alpha_0) \\ \mathbf{1}(y_i \leq \bar{q}_{0k}) \mathbf{1}(z_i = k)(1 - T_i) - \left(\frac{1 - \alpha_0}{1 - \alpha_0 - \alpha_1} \right) \mathbf{1}(z_i = k)(1 - T_i - \alpha_1) \\ \mathbf{1}(y_i \leq \underline{q}_{1k}) \mathbf{1}(z_i = k) T_i - \left(\frac{1 - \alpha_1}{1 - \alpha_0 - \alpha_1} \right) \mathbf{1}(z_i = k)(T_i - \alpha_0) \\ \mathbf{1}(y_i \leq \bar{q}_{1k}) \mathbf{1}(z_i = k) T_i - \left(\frac{\alpha_0}{1 - \alpha_0 - \alpha_1} \right) \mathbf{1}(z_i = k)(1 - T_i - \alpha_1) \end{bmatrix}$$