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

Francis J. DiTraglia
Camilo García-Jimeno

University of Pennsylvania

October 11th, 2018

Binary Regressors

- Thank you for inviting me. Joint work with Camilo Garcia-Jimeno.
- Intro. 'metrics students learn that a valid IV serves double duty: correct for endogeneity and classical measurement error
- Classical measurement error is a special case: requires true value of regressor indep. of or at least uncorrelated with measurement error
- Applied work often involves endogenous binary regressor: smoker/non-smoker or union/non-union. Binary $\implies$ non-classical error. True 0 $\implies$ can only mis-measure *upwards* as 1; true 1 $\implies$ can only mis-measure *downwards* as 0. Error *negatively correlated* with truth.
- To accommodate this, consider *non-diff* error. Say more later, but roughly non-diff means *conditionally classical*: condition on truth and controls, remaining component of error unrelated to everything else.
- Today pose simple question: binary, endog. regressor subject to non-diff. error. Can valid IV correct for *both* measurement error and endog?

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

(Additively Separable ε and binary T^* $\Rightarrow$ linear model given $\mathbf{x}$)

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

(Additively Separable ε and binary $T^* \Rightarrow$ linear model given $\mathbf{x}$)

- Here is the specific model I will focus on today. Additively separable model, want to learn the causal effect of binary regressor T^* on y. Unfortunately T^* is unobserved. Observe only mis-measured binary surrogate T. To make matters worse, T^* is endogenous, but we have a discrete instrument z.
- Additive separability is an assumption. Allow very general forms of observed heterogeneity through $\mathbf{x}$ but restricts unobserved heterogeneity.
- Conditionally linear model. This is without loss of generality since the model is additively separable and T^* is binary.
- Mainly focus on additively separable case today, but will also discuss implications of our results for a LATE model.

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

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

Constributions of This Paper

1. Show that only existing point identification result for mis-classified, endogenous T^* is incorrect.
2. Derive sharp identified set for $\beta(\mathbf{x})$ under standard assumptions.
3. Prove point identification of $\beta(\mathbf{x})$ under slightly stronger assumptions.
4. Point out problem of weak identification in mis-classification models, develop identification-robust inference for $\beta(\mathbf{x})$.

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

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

Constributions of This Paper

1. Show that only existing point identification result for mis-classified, endogenous T^* is incorrect.
2. Derive sharp identified set for $\beta(\mathbf{x})$ under standard assumptions.
3. Prove point identification of $\beta(\mathbf{x})$ under slightly stronger assumptions.
4. Point out problem of weak identification in mis-classification models, develop identification-robust inference for $\beta(\mathbf{x})$.

- Here are the main contributions of paper that I will discuss today.
- Many papers consider using IV to identify effect of exog. mis-measured binary regressor, but little work on endog. case. First: show only point identification result for this case incorrect: ident. is an open question.
- Next: use standard assumptions to derive the "sharp identified set" for β. This means *fully* exploit all information in the data and our assumptions to derive tightest possible bounds for β. If bounds contain a single point, β is point identified. Otherwise partially identified.
- Novel and informative bounds for β, but not point identified. Then consider slightly stronger assumptions that allow us to exploit additional features of the data and show that these suffice to point identify β.
- Next consider inference. Show that mis-classification models, suffer from potential weak identification. Propose procedure for robust inference.
- Now a motivating example…

Example: Schooling and Test Scores

Burde & Linden (2013, AEJ Applied)

RCT in Afghanistan: schools built in randomly selected villages. In treatment villages only some girls attend school; in control villages some girls attend school elsewhere.

- y – Girl's score on math and language test
- T^* – Girl's true school attendance
- T – Parent's report of child's school attendance
- $\mathbf{x}$ – Child and household characteristics
- z – School built in village

Example: Schooling and Test Scores

Burde & Linden (2013, AEJ Applied)

RCT in Afghanistan: schools built in randomly selected villages. In treatment villages only some girls attend school; in control villages some girls attend school elsewhere.

- y – Girl's score on math and language test
- T^* – Girl's true school attendance
- T – Parent's report of child's school attendance
- $\mathbf{x}$ – Child and household characteristics
- z – School built in village

PUT SOME NOTES HERE!

"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: $z \in \{0, 1\}$

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

If T^* were observed, these conditions would identify β.

"Baseline" Assumptions II – Measurement Error

Notation: Mis-classification Rates

"↑" $\alpha_0(\mathbf{x}, z) \equiv \mathbb{P}\left(T = 1 \middle| T^* = 0, \mathbf{x}, z\right)$

"↓" $\alpha_1(\mathbf{x}, z) \equiv \mathbb{P}\left(T = 0 \middle| T^* = 1, \mathbf{x}, z\right)$

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

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

Non-differential Mis-classification

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

Existing Results

Correct Result – Exogenous T^*

- Mahajan (2006) Theorem 1, Frazis & Loewenstein (2003), ...
- $\mathbb{E}[\varepsilon|\mathbf{x}, z, T^*] = 0$ + "Baseline" $\Rightarrow \beta(\mathbf{x})$ identified.

Incorrect Result – Endogenous T^*

- Mahajan (2006; Ecta) A.2
- $\mathbb{E}[\varepsilon|\mathbf{x}, z, T^*, T] = \mathbb{E}[\varepsilon|\mathbf{x}, T^*]$ + "Baseline" $\Rightarrow \beta(\mathbf{x})$ identified.

We show that Mahajan's assumptions imply that the instrument is uncorrelated with T^* unless T^* is *exogenous*.

Existing Results

Correct Result – Exogenous T^*

- Mahajan (2006) Theorem 1, Frazis & Loewenstein (2003), ...
- $\mathbb{E}[\varepsilon|\mathbf{x}, z, T^*] = 0$ + "Baseline" $\Rightarrow \beta(\mathbf{x})$ identified.

Incorrect Result – Endogenous T^*

- Mahajan (2006; Ecta) A.2
- $\mathbb{E}[\varepsilon|\mathbf{x}, z, T^*, T] = \mathbb{E}[\varepsilon|\mathbf{x}, T^*]$ + "Baseline" $\Rightarrow \beta(\mathbf{x})$ identified.

We show that Mahajan's assumptions imply that the instrument is uncorrelated with T^* unless T^* is *exogenous*.

- Point out that the FL estimator is a nonlinear GMM rather than IV and note that they require*joint* exogeneity of T^* and z.
- 1st contribution: show that only existing point identification result for mis-measured, binary, endog. regressor is false
- As mentioned a few minutes ago, main result from Mahajan (2006; Ecta) is for T^*, but paper also contains a result for the endogenous case [READ THE RESULT]
- Exotic–looking assumption is needed to leverage Mahajan's result for the exogenous case. Unfortunately we show that it leads to a contradiction. [READ THE RESULT]
- Identification in this model is an open question: though Mahajan's proof fails, this does not establish that β is unidentified under the baseline assumptions.
- Next step show you two known results: simple bounds for α_0, α_1, and relationship between IV estimator and α_0, α_1, yielding bounds for β
- Then our 2nd contribution: sharp identified set for β under baseline

Simple Bounds for Mis-classification from First-stage

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

Relationship

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

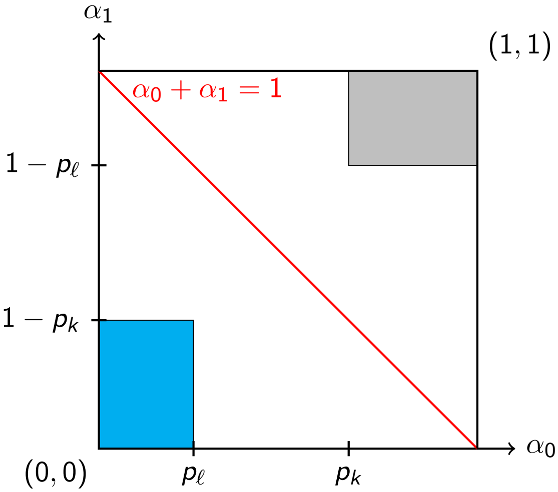
z does not affect (α_0, α_1); denominator $\neq 0$

Bounds for Mis-classification

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

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

$$\alpha_0 \leq \min_k \{p_k\}, \quad \alpha_1 \leq \min_k \{1 - p_k\}$$



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)

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

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

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

Known Result

- $\beta(\mathbf{x})$ is between Wald and Reduced form; same sign as Wald.
- Doesn't rely on non-differential assumption or additive sep.
- Frazis & Loewenstein (2003), Ura (2016), ...

Non-differential Assumption

- $\mathbb{E}[\varepsilon|\mathbf{x}, T^*, T, z] = \mathbb{E}[\varepsilon|\mathbf{x}, T^*, z]$
- Used in literature to identify $\beta(\mathbf{x})$ when T^* is exogenous.
- Does it restrict the identified set when T^* is endogenous?

Binary Regressors

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

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

Known Result

- $\beta(\mathbf{x})$ is between Wald and Reduced form; same sign as Wald.
- Doesn't rely on non-differential assumption or additive sep.
- Frazis & Loewenstein (2003), Ura (2016), . . .

Non-differential Assumption

- $\mathbb{E}[\varepsilon|\mathbf{x}, T^*, T, z] = \mathbb{E}[\varepsilon|\mathbf{x}, T^*, z]$
- Used in literature to identify $\beta(\mathbf{x})$ when T^* is exogenous.
- Does it restrict the identified set when T^* is endogenous?

Answer turns out to be yes. Now talk about our 2nd contribution: deriving the so-called "sharp" identified set

Restrictions from Non-differential Mis-classification?

(Suppress **x** for simplicity)

Notation

- $r_{tk} \equiv \mathbb{P}(T^* = 1 | T = t, z = k)$, function of (α_0, α_1) and observables only
- 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 **x** for simplicity)

Notation

- $r_{tk} \equiv \mathbb{P}(T^* = 1 | T = t, z = k)$, function of (α_0, α_1) and observables only
- z_k is shorthand for $z = k$

Adding Non-differential Assumption

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

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

$\boxed{\text{2 equations in 2 unknowns} \Rightarrow \text{solve for } \mathbb{E}(y|T^* = t^*, z = k) \text{ given } (\alpha_0, \alpha_1).}$

Restrictions from Non-differential Mis-classification?

Law of Total Probability

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

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

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

Previous Slide

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

Key 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 $\mu = \text{mean}(H)$?

Valid CDFS

$$
\begin{aligned}
&0 \leq H \leq 1 \\
&0 \leq G \leq 1 \quad \iff \quad [F - (1-p)]/p \leq H \leq F/p
\end{aligned}
$$

$$
\boxed{\max\left\{0,\ \frac{F(x)}{p} - \frac{1-p}{p}\right\} \leq H(x) \leq \min\left\{1, \frac{F(x)}{p}\right\}}
$$

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 $\mu = \text{mean}(H)$?

Valid CDFS

$$\begin{array}{l} 0 \leq H \leq 1 \\ 0 \leq G \leq 1 \quad \iff \quad [F - (1-p)]/p \leq H \leq F/p \end{array}$$

$$\boxed{\max\left\{0, \frac{F(x)}{p} - \frac{1-p}{p}\right\} \leq H(x) \leq \min\left\{1, \frac{F(x)}{p}\right\}}$$

Explain how we solve to eliminate G and get conditions on H only!

Restrictions from Non-differential Mis-classification?

Notation

$$\overline{H} \equiv \max\left\{0,\ \frac{F(x)}{p} - \frac{1-p}{p}\right\}, \quad \underline{H} \equiv \min\left\{1, \frac{F(x)}{p}\right\}$$

1st Order Stochastic Dominance

$$\overline{H}(x) \leq H(x) \leq \underline{H}(x) \quad \text{for all } x$$

$$\implies \underbrace{\int_{\mathbb{R}} x\, \underline{H}(dx)}_{\underline{\mu}(p,F)} \leq \underbrace{\int_{\mathbb{R}} x\, H(dx)}_{\mu} \leq \underbrace{\int_{\mathbb{R}} x\, \overline{H}(dx)}_{\overline{\mu}(p,F)}$$

Restrictions from Non-differential Mis-classification?

Notation

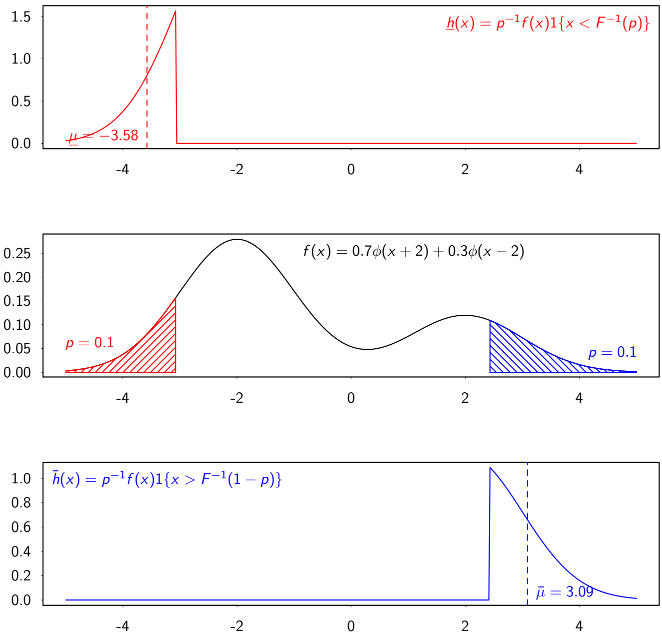
$$\overline{H} \equiv \max\left\{0, \frac{F(x)}{p} - \frac{1-p}{p}\right\}, \quad \underline{H} \equiv \min\left\{1, \frac{F(x)}{p}\right\}$$

1st Order Stochastic Dominance

$$\overline{H}(x) \leq H(x) \leq \underline{H}(x) \quad \text{for all } x$$

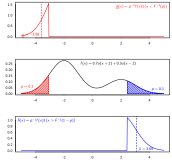
$$\implies \underbrace{\int_{\mathbb{R}} x\, \underline{H}(dx)}_{\underline{\mu}(p,F)} \leq \underbrace{\int_{\mathbb{R}} x\, H(dx)}_{\mu} \leq \underbrace{\int_{\mathbb{R}} x\, \overline{H}(dx)}_{\overline{\mu}(p,F)}$$

Explain the somewhat odd-looking convention for $\underline{H}$ and $\overline{H}$. Constraint from preceding slide is a stochastic dominance relationship. It implies an inequality for means.

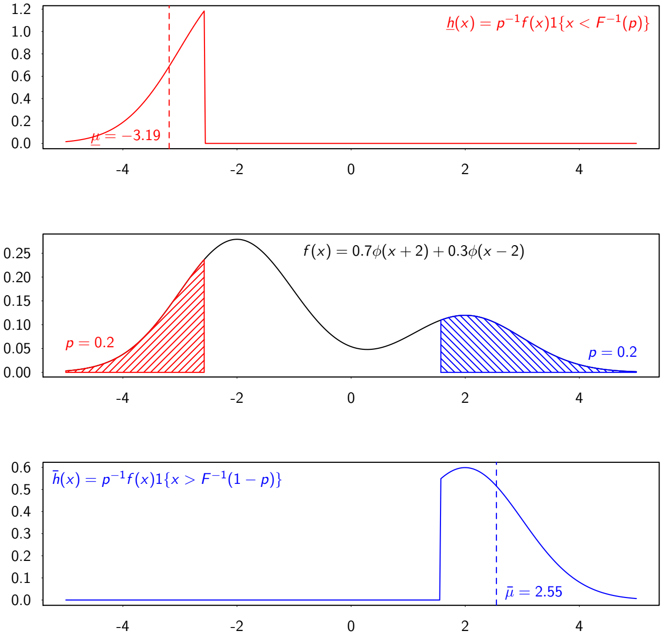


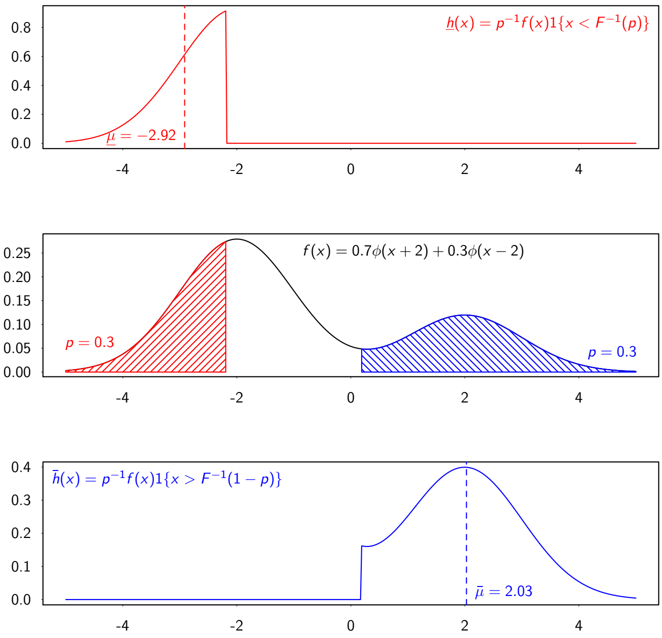
$\underline{h}(x) = p^{-1} f(x) 1\{x < F^{-1}(p)\}$
$\underline{\mu} = -3.58$
$f(x) = 0.7\phi(x+2) + 0.3\phi(x-2)$
$p = 0.1$
$p = 0.1$
$\bar{h}(x) = p^{-1} f(x) 1\{x > F^{-1}(1-p)\}$
$\bar{\mu} = 3.09$

Binary Regressors

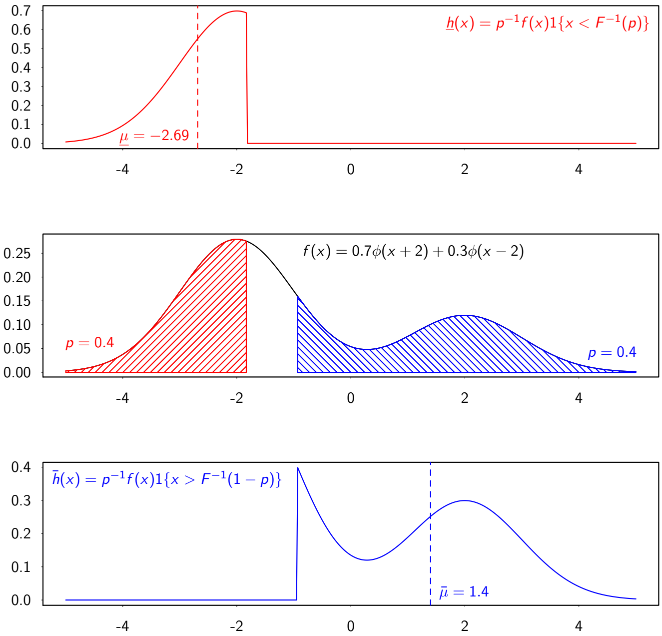


Picture very simple: for given weight p on H, top panel shows the smallest mean that H can have and the bottom shows the largest mean it can take to yield a valid mixture in which H has weight p. As you change p, you change the range of values that the mean of H can take. In this example the observed distribution F is a simple mixture of normals. If it were a different distribution we'd get different restrictions: picture shows how shape of F leads to the upper and lower bounds for μ.

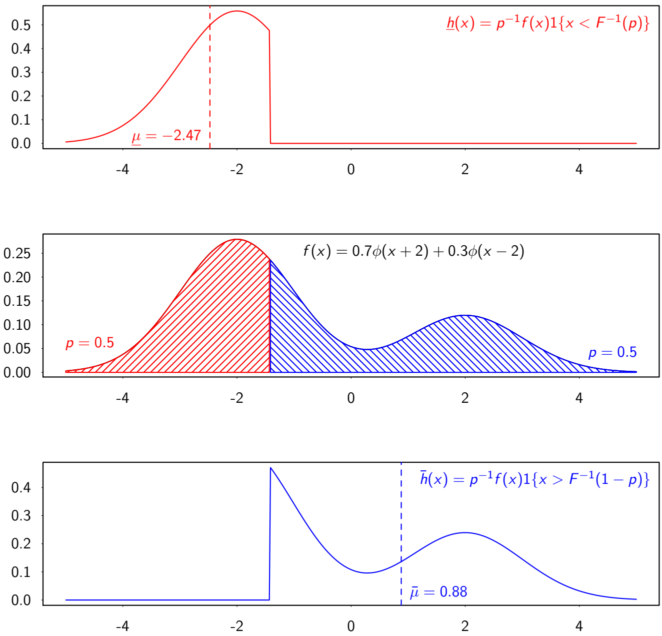

$\underline{h}(x) = p^{-1}f(x)1\{x < F^{-1}(p)\}$
$\underline{\mu} = -3.19$
$f(x) = 0.7\phi(x+2) + 0.3\phi(x-2)$
$p = 0.2$
$p = 0.2$
$\bar{h}(x) = p^{-1}f(x)1\{x > F^{-1}(1-p)\}$
$\bar{\mu} = 2.55$



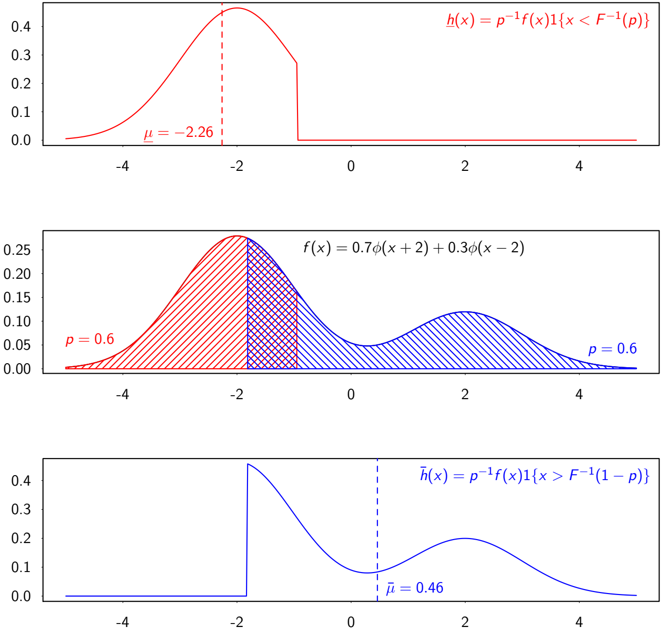
$\underline{h}(x) = p^{-1} f(x) 1\{x < F^{-1}(p)\}$
$\underline{\mu} = -2.92$
$f(x) = 0.7\phi(x+2) + 0.3\phi(x-2)$
$p = 0.3$
$p = 0.3$
$\bar{h}(x) = p^{-1} f(x) 1\{x > F^{-1}(1-p)\}$
$\bar{\mu} = 2.03$



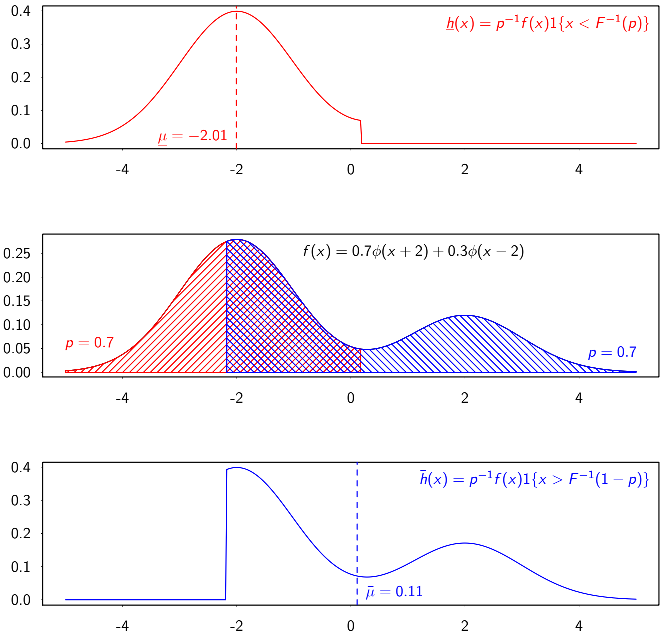
$\underline{h}(x) = p^{-1}f(x)1\{x < F^{-1}(p)\}$
$\underline{\mu} = -2.69$
$f(x) = 0.7\phi(x+2) + 0.3\phi(x-2)$
$p = 0.4$
$p = 0.4$
$\bar{h}(x) = p^{-1}f(x)1\{x > F^{-1}(1-p)\}$
$\bar{\mu} = 1.4$

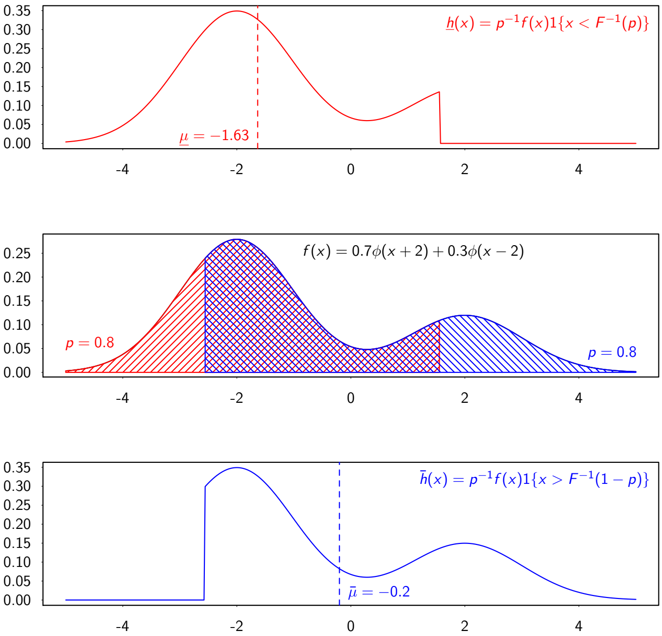


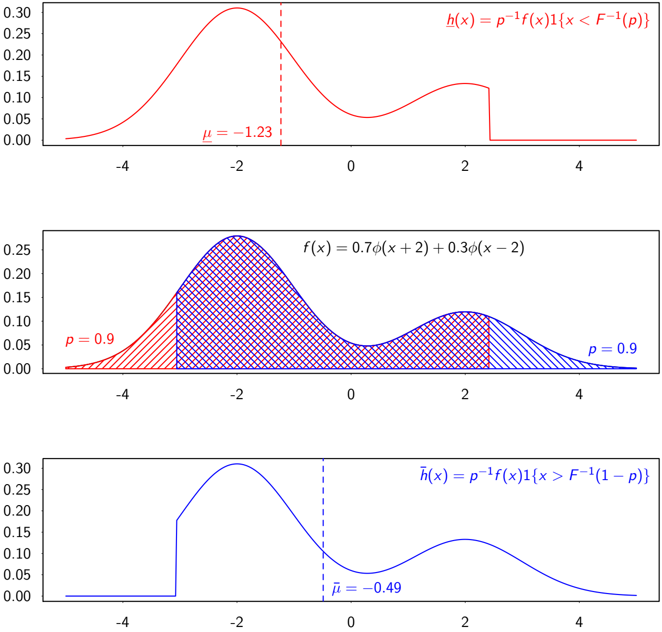
$\underline{h}(x) = p^{-1}f(x)1\{x < F^{-1}(p)\}$
$\underline{\mu} = -2.47$
$f(x) = 0.7\phi(x+2) + 0.3\phi(x-2)$
$p = 0.5$
$p = 0.5$
$\bar{h}(x) = p^{-1}f(x)1\{x > F^{-1}(1-p)\}$
$\bar{\mu} = 0.88$

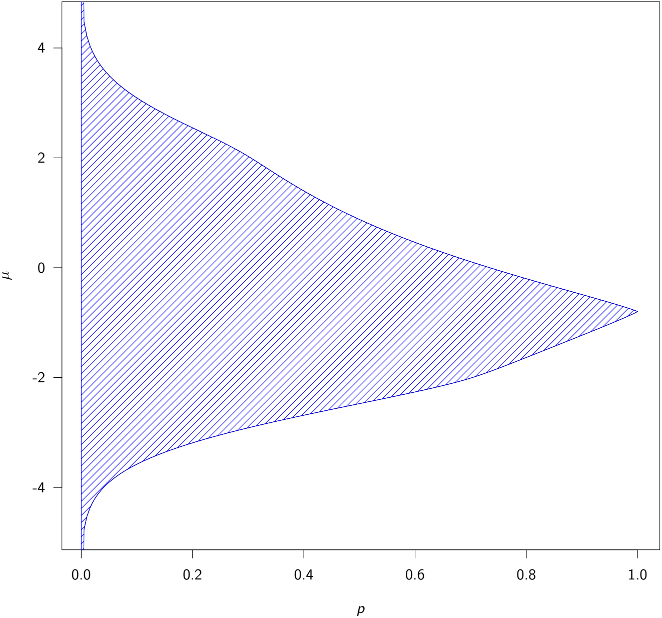


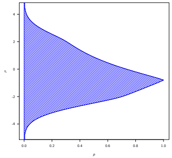
$\underline{h}(x) = p^{-1} f(x) 1\{x < F^{-1}(p)\}$
$\underline{\mu} = -2.26$
0.4
0.3
0.2
0.1
0.0
-4
-2
0
2
4
$f(x) = 0.7\phi(x+2) + 0.3\phi(x-2)$
$p = 0.6$
$p = 0.6$
0.25
0.20
0.15
0.10
0.05
0.00
$\bar{h}(x) = p^{-1} f(x) 1\{x > F^{-1}(1-p)\}$
$\bar{\mu} = 0.46$


$\underline{h}(x) = p^{-1}f(x)1\{x < F^{-1}(p)\}$
$\underline{\mu} = -2.01$
0.4
0.3
0.2
0.1
0.0
-4
-2
0
2
4
$f(x) = 0.7\phi(x+2) + 0.3\phi(x-2)$
$p = 0.7$
$p = 0.7$
0.25
0.20
0.15
0.10
0.05
0.00
-4
-2
0
2
4
$\bar{h}(x) = p^{-1}f(x)1\{x > F^{-1}(1-p)\}$
$\bar{\mu} = 0.11$
0.4
0.3
0.2
0.1
0.0
-4
-2
0
2
4


$\underline{h}(x) = p^{-1} f(x) 1\{x < F^{-1}(p)\}$
$\underline{\mu} = -1.63$
$f(x) = 0.7\phi(x+2) + 0.3\phi(x-2)$
$p = 0.8$
$p = 0.8$
$\bar{h}(x) = p^{-1} f(x) 1\{x > F^{-1}(1-p)\}$
$\bar{\mu} = -0.2$


$\underline{h}(x) = p^{-1}f(x)1\{x < F^{-1}(p)\}$
$\underline{\mu} = -1.23$
$f(x) = 0.7\phi(x+2) + 0.3\phi(x-2)$
$p = 0.9$
$p = 0.9$
$\bar{h}(x) = p^{-1}f(x)1\{x > F^{-1}(1-p)\}$
$\bar{\mu} = -0.49$


4
2
0
-2
-4
μ
0.0
0.2
0.4
0.6
0.8
1.0
p



As p approaches 1, the mean of H is more tightly constrained: must be close to the mean of the observed distribution F, namely -0.8. As p approaches zero, it is less and less constrained: since it contributes very little to the overall mixture, it can take on nearly any mean.

Restrictions from Non-differential Mis-classification

Necessary and Sufficient Condition if F is Continuous

$$\int_{-\infty}^{F^{-1}(p)} \frac{x}{p}\, f(x)\, dx \leq \ \mu \ \leq \int_{F^{-1}(1-p)}^{+\infty} \frac{x}{p}\, f(x)\, dx$$

Back to Our Original Problem

- Observe F_{tk} for all (t, k)
- r_{tk} pinned down by (α_0, α_1)
- Can we find $F_{tk}^{t^*}$ so that $F_{tk} = (1 - r_{tk})F_{tk}^0 + r_{tk}F_{tk}^1$?
- Non-diff. assumption $\Rightarrow$ mean of F_{tk}^1 pinned down by (α_0, α_1).
- Implies joint restrictions on (α_0, α_1), hence β.

Restrictions from Non-differential Mis-classification

Necessary and Sufficient Condition if F is Continuous

$$\int_{-\infty}^{F^{-1}(p)} \frac{x}{p}\, f(x)\, dx \leq \mu \leq \int_{F^{-1}(1-p)}^{+\infty} \frac{x}{p}\, f(x)\, dx$$

Back to Our Original Problem

- Observe F_{tk} for all (t, k)
- r_{tk} pinned down by (α_0, α_1)
- Can we find $F_{tk}^{t^*}$ so that $F_{tk} = (1 - r_{tk})F_{tk}^0 + r_{tk}F_{tk}^1$?
- Non-diff. assumption $\Rightarrow$ mean of F_{tk}^1 pinned down by (α_0, α_1).
- Implies joint restrictions on (α_0, α_1), hence β.

I just showed you necessary conditions, but it turns out that they're also sufficient if F is continuous, so we've solved the problem of when a valid mixture decomposition exists. How does this relate to the original question: does non-differential measurement error further restrict α_0, α_1?

Sharp Identified Set under Baseline Assumptions

Theorem

Under baseline assumptions, sharp identified set for $\beta(\mathbf{x})$ is never a singleton, regardless of how many (discrete) values z takes on.

Intuition

No mis-classification $\Rightarrow r_{tk} = 0$ or 1 and we can always form a valid mixture in this case. Show that Wald estimand always lies within the sharp identified set for β.

Theorem

Under baseline assumptions, sharp identified set for $\beta(\mathbf{x})$ is never a singleton, regardless of how many (discrete) values z takes on.

Intuition

No mis-classification $\Rightarrow r_{tk} = 0$ or 1 and we can always form a valid mixture in this case. Show that Wald estimand always lies within the sharp identified set for β.

Second contribution. Simple bounds I showed you earlier are not sharp: in other words, they're not the best bounds you can get under our assumptions. Even when we get the best bounds (the "sharp" bounds) they're not enough to point identify β. Also want to point out that the restrictions from non-differential measurement error can be very informative in practice! Now transition to point identification argument. Can we obtain point identification under stronger but credible assumptions?

Point Identification: 1st Ingredient

Reparameterization

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

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

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

$$\boxed{\beta(\mathbf{x}) = 0 \iff \theta_1(\mathbf{x}) = \theta_2(\mathbf{x}) = \theta_3(\mathbf{x}) = 0}$$

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

(Baseline) + (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

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

Hence, $\beta(\mathbf{x})$ is identified if mis-classification is one-sided.

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) $\implies$

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

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 Conditions for (II) and (III)

(i) T is conditionally independent of (ε, z) given $(T^*, \mathbf{x})$

(ii) z is conditionally independent of ε given $\mathbf{x}$

Just-Identified System of Moment Equalities

Suppress dependence on $\mathbf{x}$ to simplify the notation from here on...

Collect Lemmas from Above:

$$
\begin{aligned}
\mathrm{Cov}(y, z) - \mathrm{Cov}(T, z)\theta_1 &= 0 \\
\mathrm{Cov}(y^2, z) - 2\mathrm{Cov}(yT, z)\theta_1 + \mathrm{Cov}(T, z)\theta_2 &= 0 \\
\mathrm{Cov}(y^3, z) - 3\mathrm{Cov}(y^2 T, z)\theta_1 + 3\mathrm{Cov}(yT, z)\theta_2 - \mathrm{Cov}(T, z)\theta_3 &= 0
\end{aligned}
$$

Notation: Observed Data Vector

$$
\mathbf{w}_i' = (T_i, y_i, y_i T_i, y_i^2, y_i^2 T_i, y_i^3)
$$

Just-Identified System of Moment Equalities

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

$$\boldsymbol{\Psi} = \begin{bmatrix} \psi_1 & \psi_2 & \psi_3 \end{bmatrix} \qquad \boldsymbol{\kappa} = (\kappa_1, \kappa_2, \kappa_3)' \equiv \text{ ``Intercepts''}$$

$$\psi_1' = \begin{bmatrix} -\theta_1 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad \theta_1 = \beta/(1 - \alpha_0 - \alpha_1)$$

$$\psi_2' = \begin{bmatrix} \theta_2 & 0 & -2\theta_1 & 1 & 0 & 0 \end{bmatrix} \qquad \theta_2 = \theta_1^2\left[1 + \alpha_0 - \alpha_1\right]$$

$$\psi_3' = \begin{bmatrix} -\theta_3 & 0 & 3\theta_2 & 0 & -3\theta_1 & 1 \end{bmatrix} \qquad \theta_3 = \theta_1^3\left[(1 - \alpha_0 - \alpha_1)^2 + 6\alpha_0(1 - \alpha_1)\right]$$

Weak Identification Problem

Moment are conditions uninformative about (α_0, α_1) when β is small: GMM performs very badly. ▸ sims

Inference for a Mis-classified Regressor

The Problem

- 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 Solution

- Sharp identified set result from above remains informative even if β is small or zero.
- Implies a number of *inequality* moment conditions
- Identification-robust inference combining equality and inequality moment conditions based on generalized moment selection (GMS)

Moment Inequalities I – First-stage Probabilities

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

$$m(\mathbf{w}_i, \vartheta) \equiv \left[\begin{array}{l} \mathbf{1}(z_i = k)(T - \alpha_0) \\ \mathbf{1}(z_i = k)(1 - T_i - \alpha_1) \end{array} \right]$$

Moment Inequalities II – Non-differential Assumption

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

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

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

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

Inference with Moment Equalities and Inequalities

Moment Conditions

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

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

Test Statistic

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

Critical Value

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

Inference with Moment Equalities and Inequalities

Moment Conditions

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

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

Test Statistic

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

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 $H_0\colon \vartheta = \vartheta_0$

Explain about the meaning of the m-var, the sigma-hat and the "minus" subscript

Inference with Moment Equalities and Inequalities

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.
- Drop inequality j if $\frac{\sqrt{n}\,\bar{m}_{n,j}(\vartheta_0)}{\widehat{\sigma}_{n,j}(\vartheta_0)} > \sqrt{\log n}$
- Uniformly valid test of $H_0\colon \vartheta = \vartheta_0$ 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

Leverage Special Structure of Model

- β only enters MCs through $\theta_1 = \beta/(1 - \alpha_0 - \alpha_1)$
- If z is a strong instrument, inference for θ_1 is standard.
- $(\boldsymbol{\kappa}, \mathbf{q})$ strongly identified under null for (α_0, α_1)

Procedure

1. Concentrate out $(\theta_1, \boldsymbol{\kappa}, \boldsymbol{q}) \Rightarrow$ 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 $\Rightarrow (1 - \delta - \delta_2) \times 100\%$ CI for β

Our Solution: Bonferroni-Based Inference

Leverage Special Structure of Model

- β only enters MCs through $\theta_1 = \beta/(1-\alpha_0-\alpha_1)$
- If z is a strong instrument, inference for θ_1 is standard.
- $(\boldsymbol{\kappa}, \mathbf{q})$ strongly identified under null for (α_0, α_1)

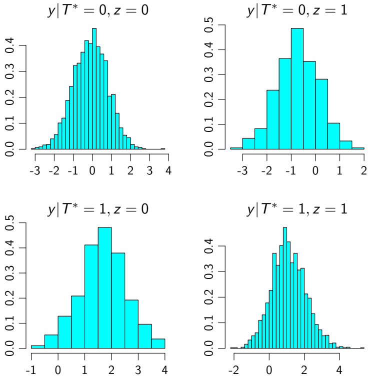
Procedure

1. Concentrate out $(\theta_1, \boldsymbol{\kappa}, \boldsymbol{q}) \Rightarrow$ 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 $\Rightarrow (1-\delta-\delta_2) \times 100\%$ CI for β

Explain that the procedure works well in simulations etc. Possibly add link to simulation here.

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

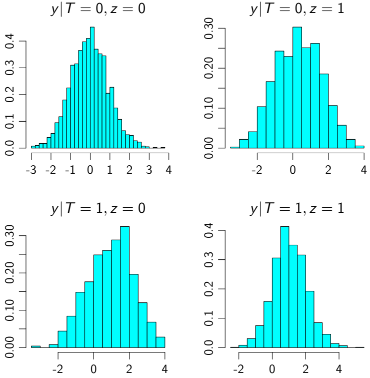
Results if T^* were observed



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

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

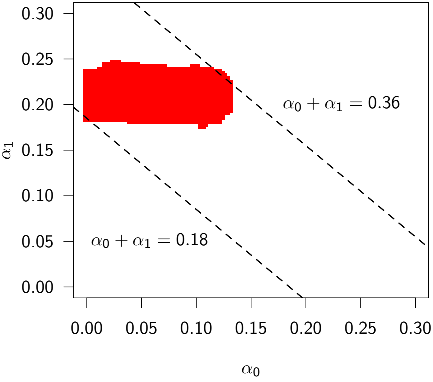
Results using T instead of T^*



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

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)



Bonferroni Interval

1. 97.5% CI for $(1 - \alpha_0 - \alpha_1) = (0.64, 0.82)$
2. 97.5% CI for $\theta_1 = (1.20, 1.47)$
3. $>$ 95% CI for β:
   $(0.64 \times 1.20, 0.82 \times 1.47) = (0.77, 1.21)$

Comparisons

- (0.88, 1.04) for IV if T^* were observed
- (1.22,1.45) for naive IV interval using T

Conclusion

- Identification and inference for effect of binary, mis-classified, endogenous regressor.
- Show that only existing point identification result is incorrect.
- Derive sharp identified set for $\beta(\mathbf{x})$ under standard assumptions.
- Prove point identification of $\beta(\mathbf{x})$ under slightly stronger assumptions.
- Point out problem of weak identification in mis-classification models, develop identification-robust inference for $\beta(\mathbf{x})$.