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

Francis J. DiTraglia

Camilo García-Jimeno

University of Pennsylvania

September 14th, 2017

Additively Separable Model

$$y = h(T^*, \mathbf{x}) + \varepsilon$$

- y – Outcome of interest
- h – Known or unknown function
- T^* – Unobserved, endogenous binary regressor
- T – Observed, mis-measured binary surrogate for T^*
- $\mathbf{x}$ – Exogenous covariates
- ε – Mean-zero error term

What is the Effect of T^*?

Re-write the Model

$$y = c(\mathbf{x}) + \beta(\mathbf{x})T^* + \varepsilon$$
$$\beta(\mathbf{x}) = h(1, \mathbf{x}) - h(0, \mathbf{x})$$
$$c(\mathbf{x}) = h(0, \mathbf{x})$$

This Paper:

- Does a discrete instrument z (typically binary) identify $\beta(\mathbf{x})$?
- What assumptions are required for z and the surrogate T?
- How to carry out inference for a mis-classified regressor?

Example: Job Training Partnership Act (JTPA)

Heckman et al. (2000, QJE)

Randomized offer of job training, but about 30% of those *not* offered also obtain training and about 40% of those offered training don't attend. Estimate causal effect of *training* rather than *offer* of training.

- y – Log wage
- T^* – True training attendence
- T – Self-reported training attendance
- $\mathbf{x}$ – Individual characteristics
- z – Offer of job training

Related Literature

Continuous Regressor

Lewbel (1997, 2012), Schennach (2004, 2007), Chen et al. (2005), Hu & Schennach (2008), Song (2015), Hu et al. (2015)...

Binary, Exogenous Regressor

Aigner (1973), Bollinger (1996), Kane et al. (1999), Black et al. (2000), Frazis & Loewenstein (2003), Mahajan (2006), Lewbel (2007), Hu (2008)

Binary, Endogenous Regressor

Mahajan (2006), Shiu (2015), Ura (2015), Denteh et al. (2016)

"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

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

Identification Results from the Literature

Mahajan (2006) Theorem 1, Frazis & Loewenstein (2003)

$\mathbb{E}[\varepsilon|\mathbf{x}, z, T^*] = 0$, plus "Baseline" $\implies \beta(\mathbf{x})$ identified

Requires (T^*, z) jointly exogenous.

Mahajan (2006) A.2

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

Allows T^* endogenous, but we prove this claim is false.

Open Question

Do the baseline assumptions identify $\beta(\mathbf{x})$ when T^* is endogenous?

First-stage Probabilities & Mis-classification Bounds

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

Relationship

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

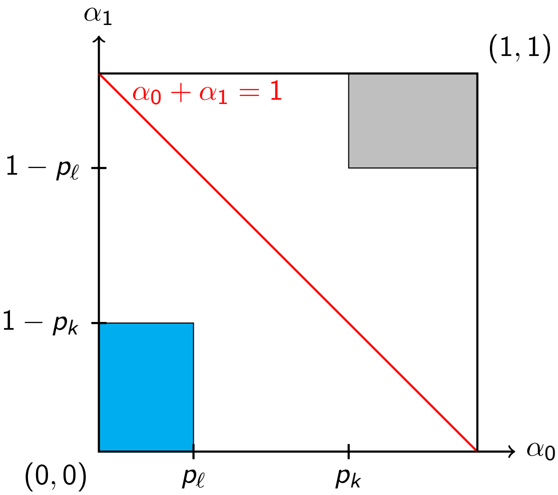
z does not affect (α_0, α_1); denominator $\neq 0$

Bounds for Mis-classification

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

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

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



Instrumental Variable Estimands

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 $\beta(\mathbf{x})$

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

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

No Mis-classification

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

Maximum Mis-classification

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

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

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

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

No Mis-classification

$\beta(\mathbf{x}) = \text{Wald}$

Maximum Mis-classification

$$\begin{aligned}\beta(\mathbf{x}) &= \text{sign}\left\{p_1(\mathbf{x}) - p_0(\mathbf{x})\right\} \times (\text{Reduced Form}) \\ &= \text{sign}\left\{\text{Wald}\right\} \times |\text{Reduced Form}|\end{aligned}$$

$\text{Wald} > 0 \iff \text{sign}\left\{p_1(\mathbf{x}) - p_0(\mathbf{x})\right\} = \text{sign}\left\{\text{Reduced Form}\right\}$

$\text{Wald} < 0 \iff \text{sign}\left\{p_1(\mathbf{x}) - p_0(\mathbf{x})\right\} \neq \text{sign}\left\{\text{Reduced Form}\right\}$

$\beta(\mathbf{x})$ has the same sign as the Wald and its magnitude is between that of Wald and Reduced Form.

Sharp Bounds?

- Bounds from the preceding slide are known in the literature.
- We prove that they are *not* sharp under the baseline assumptions from above.
- $\mathbb{E}[\varepsilon|\mathbf{x}, T^*, T, z] = \mathbb{E}[\varepsilon|\mathbf{x}, T^*, z]$ restricts (α_0, α_1) hence β.
- Description of the sharp set is somewhat complicated...
- Corollary: β is not point identified regardless of how many (discrete) values z takes on.

Point identification from slightly stronger 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 (♠)

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

Lemma

(Baseline) + (♠) $\implies$

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

Corollary

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

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

Point Identification: 1st Ingredient

Assumption (♣)

(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) + (♠) + (♣) $\implies$

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

Point Identification Result

Theorem

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

Proof Sketch

1. $\beta(\mathbf{x}) = 0 \iff \theta_1(\mathbf{x}) = 0$ so suppose this is not the case.
2. Lemmas: full-rank linear system in $\theta_1(\mathbf{x}), \theta_2(\mathbf{x}), \theta_3(\mathbf{x})$ & observables.
3. Non-linear eqs. relating $\theta_1(\mathbf{x}), \theta_2(\mathbf{x}), \theta_3(\mathbf{x})$ to $\beta(\mathbf{x})$ and $\alpha_0(\mathbf{x}), \alpha_1(\mathbf{x})$. Show that solution exists and is unique.

Sufficient Conditions for ($\spadesuit$) and ($\clubsuit$)

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

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

Just-Identified GMM Estimator: Part I

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

Collect Lemmas from Above:

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

Notation: Observed Data Vector

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

Just-Identified GMM Estimator: Part II

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

$$\boldsymbol{\Psi} = \left[\begin{array}{ccc} \boldsymbol{\psi}_1 & \boldsymbol{\psi}_2 & \boldsymbol{\psi}_3 \end{array}\right] \qquad \boldsymbol{\kappa} = (\kappa_1, \kappa_2, \kappa_3)' \equiv \text{ "Intercepts"}$$

$$\boldsymbol{\psi}_1' = \left[\begin{array}{cccccc} -\theta_1 & 1 & 0 & 0 & 0 & 0 \end{array}\right] \qquad \theta_1 = \beta/(1-\alpha_0-\alpha_1)$$

$$\boldsymbol{\psi}_2' = \left[\begin{array}{cccccc} \theta_2 & 0 & -2\theta_1 & 1 & 0 & 0 \end{array}\right] \qquad \theta_2 = \theta_1^2\left[1+\alpha_0-\alpha_1\right]$$

$$\boldsymbol{\psi}_3' = \left[\begin{array}{cccccc} -\theta_3 & 0 & 3\theta_2 & 0 & -3\theta_1 & 1 \end{array}\right] \qquad \theta_3 = \theta_1^3\left[(1-\alpha_0-\alpha_1)^2 + 6\alpha_0(1-\alpha_1)\right]$$

Weak Identification Problem

Moment conditions are uninformative about (α_0, α_1) when β is small.

Moreover, (α_0, α_1) could be on the boundary of the parameter space.

Weak Identification Problem

- β small $\Rightarrow$ moment equalities uninformative about α_1
- Same problem for other estimators from the literature but hasn't been pointed out.
- Identification robust inference: GMM Anderson-Rubin statistic
- But we can do better...
- Bounds for (α_0, α_1) immune to weak identification problem: remain informative if β is small or zero.
- Inference using Generalized Moment Selection (Andrews & Soares, 2010)

Inference With Moment Equalities and Inequalities

Moment Conditions

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

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

Test Statistic

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

$$[x]_- = \min\{x, 0\}$$

$$\bar{m}_{n,j}(\vartheta) = n^{-1} \sum_{i=1}^{n} m_j(\mathbf{w}_i, \vartheta)$$

$$\widehat{\sigma}^2_{n,j}(\vartheta) = \text{consistent est. of AVAR}\left[\sqrt{n}\ \bar{m}_{n,j}(\vartheta)\right]$$

Inference via Generalized Moment Selection

Andrews & Soares (2010)

Moment Selection Step

If $\dfrac{\sqrt{n}\,\bar{m}_{n,j}(\vartheta_0)}{\widehat{\sigma}_{n,j}(\vartheta_0)} > \sqrt{\log n}$ then drop inequality j

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 distribution of the test statistic.

Theoretical Guarantees

Uniformly valid test of $H_0\colon \vartheta = \vartheta_0$ regardless of whether ϑ_0 is identified. Not asymptotically conservative.

What we do

- Explain our multi-step procedure
- Key ingredient is the preliminary estimation of strongly identified parameters
- Details of the steps involved along with the inequalities we use and why it's ok to eliminate some parameters but not others (α_0, α_1)
- Then explain about Bonferroni and projection etc
- Projection to get inference for β, but can be conservative

Simulation DGP: $y = \beta T^* + \varepsilon$

Errors

$(\varepsilon, \eta) \sim$ jointly normal, mean 0, variance 1, correlation 0.5.

First-Stage

- Half of individuals have $z = 1$, the rest have $z = 0$.
- $T^* = \mathbf{1}\{\gamma_0 + \gamma_1 z + \eta > 0\}$
- $\delta = \mathbb{P}(T^* = 0|z = 1) = \mathbb{P}(T^* = 1|z = 0) = 0.15$

Mis-classification

- $T|T^* = 0 \sim \text{Bernoulli}(\alpha_0)$
- $T|T^* = 1 \sim \text{Bernoulli}(1 - \alpha_1)$

Simulation results!

Conclusion

- Endogenous, mis-measured binary treatment.
- Important in applied work but no solution in the literature.
- Usual (1st moment) IV assumption fails to identify β
- Higher moment / independence restrictions identify β
- Identification-Robust Inference incorportating additional inequality moment conditions.