Estimating the Effect of a Mis-measured, Endogenous, Binary Regressor

Francis J. DiTraglia

Camilo García-Jimeno

University of Pennsylvania

June 18th, 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 (JPTA)

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 Treatment

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

Binary, Exogenous Treatment

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

Binary, Endogenous Treatment

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

Baseline Assumptions – Maintained Throughout

Additively Separable Model

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

Valid & Relevant Instrument

$\mathbb{E}[\varepsilon|\mathbf{x}, z] = 0, \quad \mathbb{E}\left[T^*|\mathbf{x}, z = k\right] \neq \mathbb{E}\left[T^*|\mathbf{x}, z = \ell\right]$

Measurement Error Assumptions

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

(ii) $\alpha_0(\mathbf{x}) = \mathbb{P}(T = 1|\, T^* = 0, \mathbf{x}, z), \quad \alpha_1(\mathbf{x}) = \mathbb{P}(T = 0|\, T^* = 1, \mathbf{x}, z)$

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

Theorem

The baseline assumptions fail to identify $\beta(\mathbf{x})$, even if the instrument z takes on an arbitrarily large finite number of distinct values.

Identification from Stronger Assumptions?

Second Moment Assumption

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

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

Third Moment Assumption

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

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

Sufficient Condition

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

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

Identification Argument: Step I

Reparameterization

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

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

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

Theorem

If $\theta_1(\mathbf{x}), \theta_2(\mathbf{x})$ and $\theta_3(\mathbf{x})$ are identified and $\alpha_0(\mathbf{x}) + \alpha_1(\mathbf{x}) < 1$

- If $\theta_1(\mathbf{x}) \neq 0$, then $\beta(\mathbf{x}), \alpha_0(\mathbf{x})$ and $\alpha_1(\mathbf{x})$ are identified
- If $\theta_1(\mathbf{x}) = 0$ then $\beta(\mathbf{x})$ is identified

If $\alpha_0(\mathbf{x}) + \alpha_1(\mathbf{x}) \neq 1$, then $\beta(\mathbf{x})$ is identified up to sign.

Identification Argument: Step II

Notation

$$\pi(\mathbf{x}) = \mathrm{Cov}(T, z|\mathbf{x}), \quad \eta_j(\mathbf{x}) = \mathrm{Cov}(y^j, z|\mathbf{x}), \quad \tau_j(\mathbf{x}) = \mathrm{Cov}(Ty^j, z|\mathbf{x})$$

Theorem

Baseline plus 2nd and 3rd Moment Assumptions imply

$$\begin{aligned}
\eta_1(\mathbf{x}) &= \pi(\mathbf{x})\theta_1(\mathbf{x}) \\
\eta_2(\mathbf{x}) &= 2\tau_1(\mathbf{x})\theta_1(\mathbf{x}) - \pi(\mathbf{x})\theta_2(\mathbf{x}) \\
\eta_3(\mathbf{x}) &= 3\tau_2(\mathbf{x})\theta_1(\mathbf{x}) - 3\tau_1(\mathbf{x})\theta_2(\mathbf{x}) + \pi(\mathbf{x})\theta_3(\mathbf{x})
\end{aligned}$$

so $\theta_1(\mathbf{x}), \theta_2(\mathbf{x})$ and $\theta_3(\mathbf{x})$ are identified if $\pi(\mathbf{x}) \neq 0$.

Simple Special Case

Suppose $\alpha_0 = 0$ and No Covariates

$$\text{Cov}(y, z) - \left(\frac{\beta}{1-\alpha_1}\right)\text{Cov}(T, z) = 0$$

$$\text{Cov}(y^2, z) - \frac{\beta}{1-\alpha_1}\left\{2\text{Cov}(yT, z) - \beta\text{Cov}(T, z)\right\} = 0$$

Closed-Form Solution for β

$$\beta = \frac{2\text{Cov}(yT, z)}{\text{Cov}(T, z)} - \frac{\text{Cov}(y^2, z)}{\text{Cov}(y, z)}$$

Unconditional Moment Equalities ($\alpha_0 = 0$, No Covariates)

$$\mathbf{u}_i(\boldsymbol{\kappa}, \boldsymbol{\theta}) = \begin{bmatrix} y_i - \kappa_1 - \theta_1 T_i \\ y_i^2 - \kappa_2 - \theta_1 2 y_i T_i + \theta_2 T_i \end{bmatrix}, \quad \mathbb{E}\begin{bmatrix} \boldsymbol{u}_i(\boldsymbol{\kappa}, \boldsymbol{\theta}) \\ \mathbf{u}_i(\boldsymbol{\kappa}, \boldsymbol{\theta}) z_i \end{bmatrix} = \mathbf{0}$$

$$\begin{aligned} \theta_1 &= \beta/(1-\alpha_1) \\ \theta_2 &= \beta^2/(1-\alpha_1) \\ \kappa_1 &= c \\ \kappa_2 &= c^2 + \sigma_\varepsilon^2 \end{aligned}$$

What happens if we try standard GMM inference?

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 $z = 0$.
- $T^* = \mathbf{1}\{\gamma_0 + \gamma_1 z + \eta > 0\}$
- $\delta = \mathbb{P}(T^* = 0|z = 1) = \mathbb{P}(T^* = 1|z = 0) = 0.15$

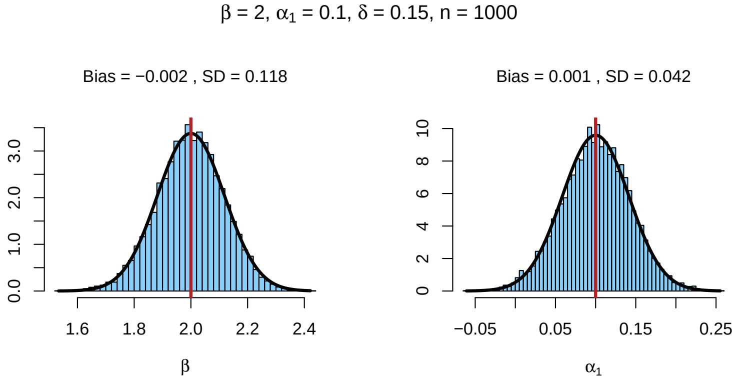
Mis-classification

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

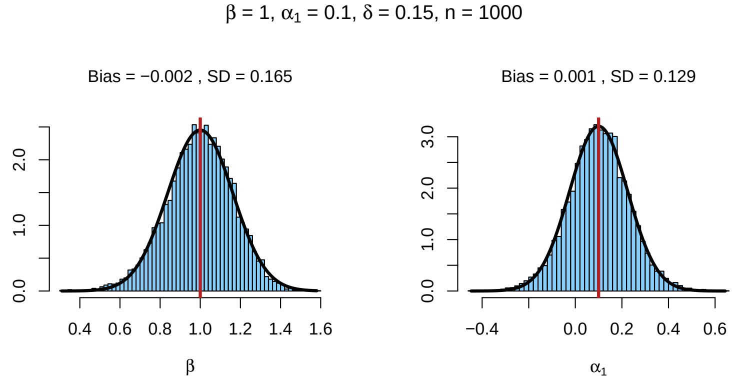
Coverage and Width of Nominal 95% GMM CIs

$\alpha_1 = 0.1, \delta = 0.15, n = 1000, \rho = 0.5$, 5000 simulation replications

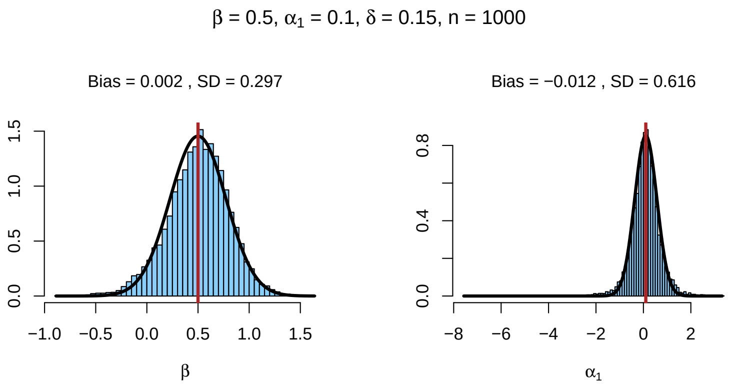
β	Coverage	Median Width
2.00	0.95	0.23
1.50	0.95	0.26
1.00	0.95	0.32
0.50	0.96	0.55
0.25	0.98	1.08
0.20	0.99	1.40
0.15	0.99	1.86
0.10	1.00	3.04
0.05	1.00	4.76
0.01	1.00	5.92



β = 2, α1 = 0.1, δ = 0.15, n = 1000
Bias = −0.002 , SD = 0.118
Bias = 0.001 , SD = 0.042
0.0
1.0
2.0
3.0
1.6
1.8
2.0
2.2
2.4
β
0
2
4
6
8
10
−0.05
0.05
0.15
0.25
α1



β = 1, α1 = 0.1, δ = 0.15, n = 1000
Bias = −0.002 , SD = 0.165
Bias = 0.001 , SD = 0.129
0.0
1.0
2.0
3.0
0.4
0.6
0.8
1.0
1.2
1.4
1.6
−0.4
0.0
0.2
0.4
0.6
β
α1



β = 0.5, α1 = 0.1, δ = 0.15, n = 1000
Bias = 0.002 , SD = 0.297
Bias = −0.012 , SD = 0.616
0.0
0.5
1.0
1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
β
0.0
0.4
0.8
−8
−6
−4
−2
0
2
α1

Weak Identification Problem

Illustrated for $\alpha_0 = 0$ but holds generally

$$\mathbf{u}_i(\boldsymbol{\kappa}, \boldsymbol{\theta}) = \begin{bmatrix} y_i - \kappa_1 - \theta_1 T_i \\ y_i^2 - \kappa_2 - \theta_1 2 y_i T_i + \theta_2 T_i \end{bmatrix}, \quad \mathbb{E}\begin{bmatrix} \boldsymbol{u}_i(\boldsymbol{\kappa}, \boldsymbol{\theta}) \\ \mathbf{u}_i(\boldsymbol{\kappa}, \boldsymbol{\theta}) z_i \end{bmatrix} = \mathbf{0}$$

$$\theta_1 = \beta/(1-\alpha_1), \quad \theta_2 = \beta^2/(1-\alpha_1)$$

- 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...

“Weak” Bounds for α_0, α_1

General Case $\alpha_0 \neq 0$

Law of Total Probability

$$p_k^* = \frac{p_k - \alpha_0}{1 - \alpha_0 - \alpha_1}, \quad 1 - p_k^* = \frac{1 - p_k - \alpha_1}{1 - \alpha_0 - \alpha_1}$$

where $p_k = \mathbb{P}(T = 1|z = k), \quad p_k^* = \mathbb{P}(T^* = 1|z = k)$

$Cor(T, T^*) > 0$

$\iff \alpha_0 + \alpha_1 < 1 \iff 1 - \alpha_0 - \alpha_1 > 0$

Implications

- $\alpha_0 < \min_k \{p_k\}, \quad \alpha_1 < \min_k \{1 - p_k\}$
- β is between β_{RF} and β_{IV}
- β_{IV} *inflated* but has correct sign

Second Moment Bounds for α_0, α_1

Observables

$$\sigma_{tk}^2 = \mathrm{Var}(y|T=t, z_k), \quad \mu_{tk} = \mathbb{E}[y|T=t, z_k], \quad p_k = \mathbb{P}(T=1|z_k)$$

Constraint on Unobservables

$$\mathrm{Var}(\varepsilon|T^* = t, z_k) > 0$$

Equivalent To

$$(p_k - \alpha_0)\left[\left(\frac{1-\alpha_0}{1-p_k}\right)\sigma_{1k}^2 - \left(\frac{\alpha_0}{p_k}\right)\sigma_{0k}^2\right] > \alpha_0(1-\alpha_0)(\mu_{1k}-\mu_{0k})^2$$

$$(1 - p_k - \alpha_1)\left[\left(\frac{1-\alpha_1}{p_k}\right)\sigma_{0k}^2 - \left(\frac{\alpha_1}{1-p_k}\right)\sigma_{1k}^2\right] > \alpha_1(1-\alpha_1)(\mu_{1k}-\mu_{0k})^2$$

Bounds can be very informative in practice...

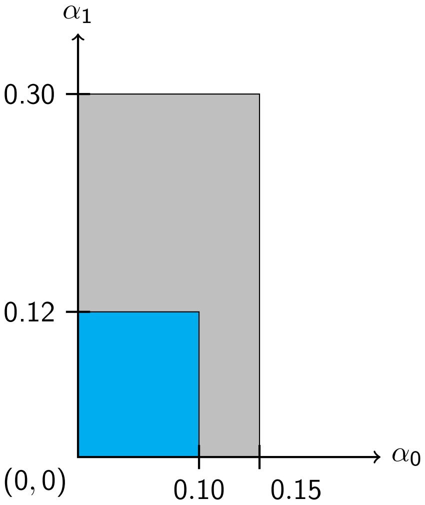
Figure based on data from Burde & Linden (2013)

"Weak" Bounds

$\beta \in [0.65 \times \beta_{IV},\ \beta_{IV}]$

Add 2nd Moments

$\beta \in [0.78 \times \beta_{IV},\ \beta_{IV}]$



Adding Auxiliary Moment Inequalities

- Bounds for (α_0, α_1) immune to weak identification problem: remain informative if β is small or zero.
- 2nd moment bounds strictly tighter, but still need weak bounds to determine which root of quadratic is extraneous.
- Since $\beta = 1/(1 - \alpha_0 - \alpha_1)$ is identified by TSLS, get meaningful restrictions on β.
- Inference using Generalized Moment Selection (Andrews & Soares, 2010)

Inference With Moment Equalities and Inequalities

Moment Conditions

$$\mathbb{E}\left[m_j(\mathbf{w}_i, \theta_0)\right] = 0, \quad j = 1, \cdots, p$$

$$\mathbb{E}\left[m_j(\mathbf{w}_i, \theta_0)\right] \geq 0, \quad j = p+1, \cdots, p+v$$

Test Statistic

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

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

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

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

Inference via Generalized Moment Selection

Andrews & Soares (2010)

Moment Selection Step

If $\dfrac{\sqrt{n}\,\bar{m}_{n,j}(\theta_0)}{\widehat{\sigma}_{n,j}(\theta_0)} > \sqrt{\ln n}$ then drop inequality j

Critical Value

- $\sqrt{n}\,\bar{m}_n(\theta_0) \to_d$ normal limit with covariance matrix $\Sigma(\theta_0)$
- Use this to bootstrap the limit distribution of the test statistic.

Theoretical Guarantees

Uniformly valid test of $H_0\colon \theta = \theta_0$ that is *not* asymptotically conservative.

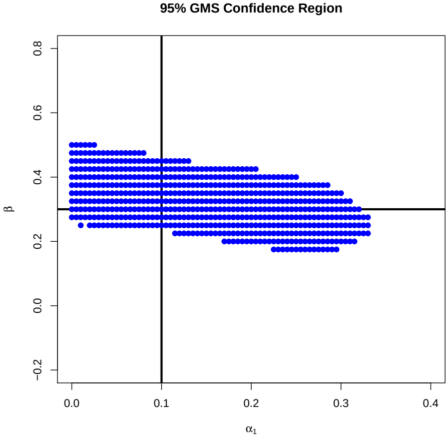
Confidence Regions

- Invert test of $\theta = \theta_0$ to form confidence region
- Preliminary estimation of strongly identified parameters ($\boldsymbol{\kappa}$)
- Yields *joint* inference for $(\alpha_0, \alpha_0, \beta)$
- Projection to get inference for β, but can be conservative

Simple Example: $n = 1000$

Simulation DGP from earlier in talk

- Special case $\alpha_0 = 1$
- $\beta = 0.25$, $\alpha_1 = 0.1$
- Reduced Form ≈ 0.18
- Wald ≈ 0.28
- Only "weak" bounds
- Naive GMM Median Width ≈ 1.1!



Additional Simulations: Size of GMS Test

Same DGP as earlier in the talk except α_0 is not zero, and use full set of inequalities...

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.