Lost in Aggregation: The Causal Interpretation of the IV estimand

Danielle Tsao, Krikamol Muandet, Frederick Eberhardt, Emilija Perković

UNIVERSITY of WASHINGTON

Introduction

Instrumental variable (IV) regression:

- *Goal: estimate average causal effect* ***(ACE)*** *of treatment A on outcome Y in presence of confounder U (top of Fig 1).*
- Regresses A and Y in two stages on a **valid** instrument I.

  → *Identifies ACE without experimental intervention.*
- IV studies in social sciences and epidemiology target effects of treatments such as ***education in years[1], GDP[2], caloric intake[3]…***

When A is an **aggregate** of component variables $A_1, \dots, A_k$ with possibly different effects on Y, we are in an *aggregate setting* as in the bottom of Fig 1.

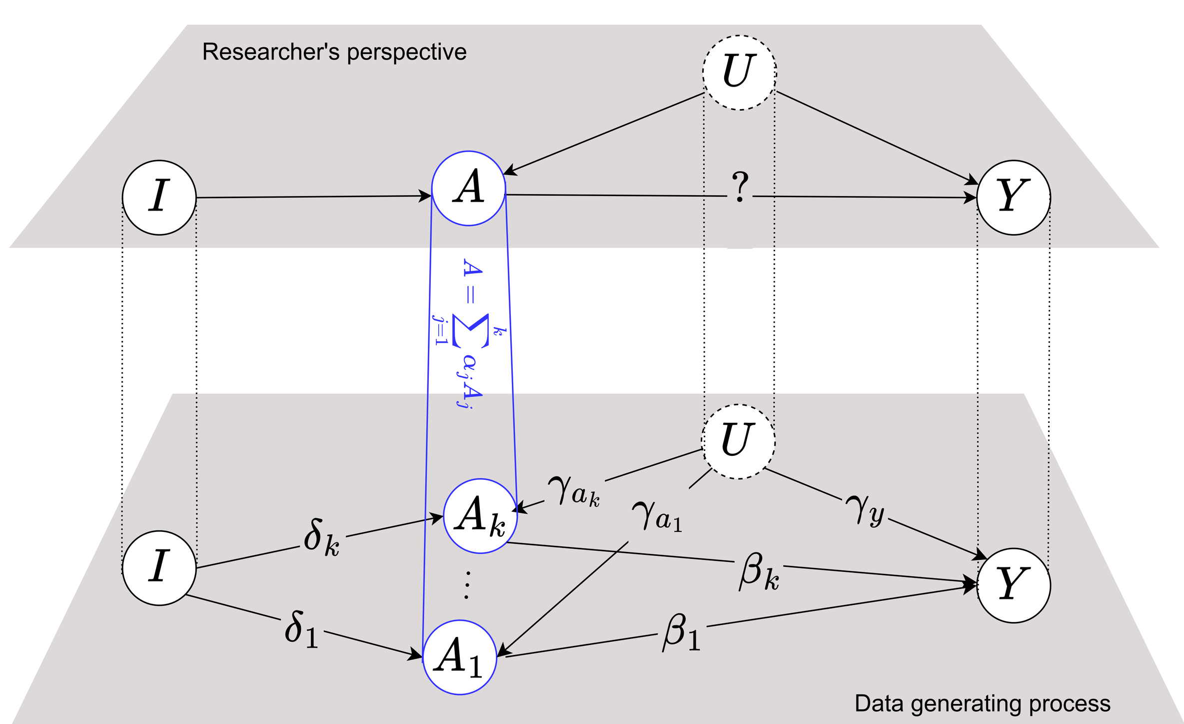


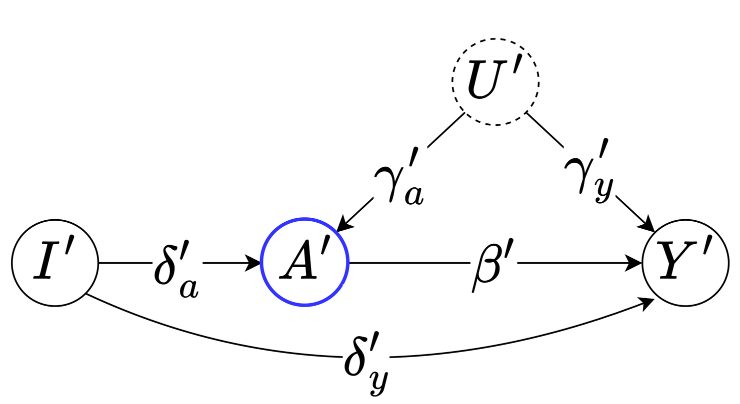
Figure 1. Classic IV setting as posited by the researcher (top) and the true aggregate setting (bottom).

Under the aggregation rule[4] $A = \sum_{j=1}^{k} \alpha_j A_j$ and the linear structural causal model corresponding to Fig 1 (bottom), the linear IV estimand, denoted β_{IV}, is given by:

$$\beta_{\mathrm{IV}} = \frac{\sum_j \beta_j \delta_j}{\sum_j \alpha_j \delta_j}. \qquad (1)$$

Q1: What is the nature of the causal effect of interest when A is an aggregate as in Fig 1?
Q2: When is this aggregate causal effect identified by the linear IV estimand in (1)?

P.S. Under Gaussianity, the aggregate setting is equivalent to the ***exclusion restriction violation*** *on the right. More on this later…*



The Aggregate Causal Effect

To answer Q1, we first consider how an intervention $do(A = a)$ is instantiated in terms of the components $A_1, \dots, A_k$.

Def. For $V = (I, U, A_1 \dots, A_k, Y)$, an interventional distribution

$$P(v)_{\mathrm{do}(A=a)} = P(i, u, a_1, \dots, a_k, y | \mathrm{do}(A = a)) \qquad (2)$$

is referred to here as a *valid aggregate-constrained component intervention distribution (ACID)* iff it satisfies the following:

i. *Surgicality* – all A_j are independent of their natural causes:
$$(2) = P(i)P(u)P^*(a_1, \dots, a_k; a)P(y \,|a_1, \dots, a_k, u).$$

ii. *Aggregation constraint* – the support of $P^*(\,\cdot\,; a)$ is:
$$\{(a_1, \dots, a_k);\ \textstyle\sum_{j=1}^{k} a_j = a\}$$

iii. *Value independence* – under the ACID, the difference
$$E[Y|\mathrm{do}(A = a + 1)] - E[Y|\mathrm{do}(A = a)] \qquad (3)$$
is independent of the value of a.

A1: *Def.* The ***aggregate causal effect*** of aggregate treatment A on an outcome Y, denoted $\mathrm{ACE}(A, Y)$, is equal to Expression (3) calculated under a valid ACID.

A valid Gaussian ACID

Consider an ACID such that under $do(A = a)$,

$$(A_1, \dots, A_k)^\top \sim N_k\left((c_1 + ad_1, \dots, c_k + ad_k)^\top, \Sigma\right),$$

where:

- For $\alpha = (\alpha_1, \dots, \alpha_k)^\top$, the *Aggregation constraint* is satisfied by

$$\underbrace{\sum_{j=1}^{k} \alpha_j c_j = 0, \quad \sum_{j=1}^{k} \alpha_j d_j = 1,}_{\text{aggregation constraint for the mean}} \quad \underbrace{\alpha^\top \Sigma = 0, \quad \Sigma \succeq 0.}_{\text{extend aggregation constraint to all realizations of } (A_1, \dots, A_k)}$$

- *Surgicality* is satisfied by construction.
- *Value independence* is satisfied as seen by the fact that

$$\mathrm{ACE}(A, Y) = \sum_{j=1}^{k} \beta_j d_j.$$

A2: The IV estimand β_{IV} identifies $\mathrm{ACE}(A, Y)$ under…

✓ ***Proportional aggregation*** *(prop. aggr.):* $\beta_j = \tau \alpha_j \ \forall j, \tau \neq 0$

✓ Or, an ***instrument-tuned intervention*** *(ITI):* e.g. in the Gaussian case, the ACID satisfies: $\sum_j \beta_j d_j = \frac{\sum_j \beta_j \delta_j}{\sum_j \alpha_j \delta_j}$.

The Exclusion Assumption

Given several instruments $I_1, \dots, I_l$ and corresponding IV estimands $\beta^1_{\mathrm{IV}}, \dots, \beta^l_{\mathrm{IV}}$, the **Sargan Test for exclusion[5]** provides partial validation for proportional aggregation.

- The **exclusion restriction** for valid instruments: $I \perp\!\!\!\perp Y | A, U$
- The Sargan Test null hypothesis is:

$$H_0\text{: for all instrument pairs } I_b, I_c \ (b, c \in \{1, \dots, l\}), \quad \beta^{I_b}_{\mathrm{IV}} = \beta^{I_c}_{\mathrm{IV}}. \qquad (4)$$

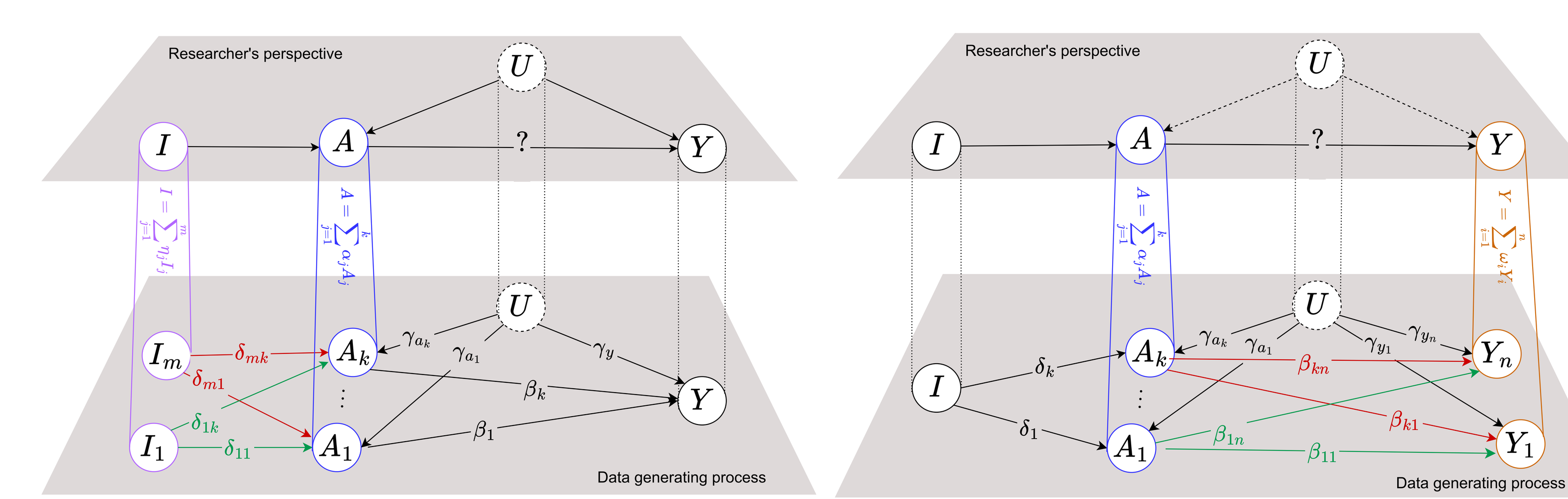
- *One* solution to (4) is proportional aggregation: $\beta_j = \tau \alpha_j \ \forall j$.

Note: If using the Sargan Test to justify IVs in an aggregate setting, watch out for the following!

- While proportional aggregation implies H_0, the converse is not true: e.g., H_0 also holds if for some κ, $\delta_{bj} = \kappa \delta_{cj} \ \forall j$.

  → *Further justification is required for prop. aggr.; i.e., why it is reasonable that the treatment effect is entirely mitigated by the aggregate.*
- The Sargan Test is known[6] to have low power for weak instruments; this problem persists in the aggregate setting.

Other aggregate settings



If the treatment is not aggregated: aggregate instruments, outcomes, or confounders pose no problems for IVs!

References

[1]Angrist and Kreuger, 1991, QJE; [2]Acemoglu et al., 2008, AER; [3]Rashad, 2006, QREF; [4]Caetano et al., 2025, preprint; [5]Windmeijer, 2019, EJ; [6]Kiviet and Kripfganz, 2021, Econ. Lett.

Acknowledgements

Many thanks to Prof. Perković, Prof. Eberhardt, and Prof. Muandet for their wonderful supervision!

McGill | Bellairs Research Institute   NSERC CRSNG   NSF