An examination of the identification problem arising from unit nonresponse in ILSA studies

IEA General Assembly, Versailles
September 2023

Diego Cortes, Jeff Dominitz, Maximiliano Romero, Sabine Meinck
A FAMILIAR EXAMPLE
A familiar example

Target inference: Was it very painful, somehow painful, or not painful?

Evidence: The picture

Missing information: The context in which he lost the tooth
A familiar example

\[ \text{truth} = f ( \text{observable, unobservable} ) \]

\[ \downarrow \]

\[ \text{conclusion} = f ( \underbrace{\text{evidence, assumption}}_{\text{observable, unobservable}} ) \]
A familiar example

1. Make no assumptions

2. Assume that my son cries when something is very painful

3. Assume the tooth fall naturally, it was loose for weeks, and the root was fully dissolved.
GENERATING KNOWLEDGE WITH IEA STUDIES
A fundamental objective of an IEA study

Generate knowledge about the distribution of student outcomes or characteristics in a given population

PIRLS    TIMSS    ICILS    ICCS
Generating knowledge
Estimation components

Statistical

Identification
THE IDENTIFICATION PROBLEM
Unit nonresponse in IEA surveys
Some insights
The basic structure

Each member of the population is characterized by:

- $y_i$ Student $i$'s CIL
- $z_i$ Student $i$'s indicator reflection participation if sampled
The basic structure

Respondents if sampled
\[(Z_i = 1)\]

\[(y_i, z_i)\]
\[(y_i, z_i)\]
\[(y_i, z_i)\]
\[(y_i, z_i)\]

Nonrespondents if sampled
\[(Z_i = 0)\]

\[(y_i, z_i)\]
\[(y_i, z_i)\]
\[(y_i, z_i)\]
\[(y_i, z_i)\]
Objective

To make conclusions about the mean of the CIL distribution in each population participating in ICILS 2018
What can be learned about the mean if all members of the population were to be sampled?
Law of Iterated Expectations

$$E[y] = E[y|z = 1]P(z = 1) + E[y|z = 0]P(z = 0)$$

Mean of $y$ among students that would participate if sampled

Mean of $y$ among students that would not participate if sampled

Proportion of students that would participate if sampled

Proportion of students that would not participate if sampled

Mean of $y$ among students that would participate if sampled

Proportion of students that would not participate if sampled
Our conclusion really depends on our choice for $E[y|z = 0]$.
Two extreme examples

- Get full certitude at the expense of credibility
- Get full credibility at the expense of certitude
Get full certitude at the expense of credibility

\[ E[y|z = 0] = E[y|z = 1] \]

Nonrespondents if sampled

\[ E[y] = E[y|z = 1] \]

Respondents if sampled
Get full certitude at the expense of credibility

Assumption:

\[ E[y|z = 0] = E[y|z = 1] \]
Get full credibility at the expense of certitude

\[ E[y|z = 0] \in [Q_5(y), Q_{95}(y)] \]

**Nonrespondents if sampled**

\[
E[y] = \left[ E[y|z = 1]P(z = 1) + Q_5(y)P(z = 0), E[y|z = 1]P(z = 1) + Q_{95}(y)P(z = 0) \right]
\]

**Respondents if sampled**
Get full credibility at the expense of certitude

Assumption:

\[ Q_5(y) \leq E[y | z = 0] \leq Q_{95}(y) \]
Two extreme examples

Between these two extreme examples there is a continuum of options for $E[y|z = 0]$. Our nonresponse model lies within this continuum.
Take aways from the exercise

1. If $P(z = 0) > 0$ then we **must** make a **choice** about $E[y|z = 0]$

2. Our choice impacts where we stand in the balance between **certitude** and **credibility**

3. If $P(z = 0) > 0$, our **conclusion** about our parameter of interest is a **weighted average** of **evidence** and assumptions
Wrap up

• Estimation framework: Partial identification
  – Establishing the boundaries of what the data reveals

• Vocal about identification problems in ILSA context
  – Assumptions matter

• Unit nonresponse
Thank you 😊

Diego.cortes@iea-hamburg.de