Estimating SDG 4 Indicators with a Bayesian Modelling Framework: Completion and Out-of-School Rates

Ameer Dharamshi
November 22th, 2022
Overview

1. Context
2. Completion Rates
3. Out-of-School Rates
   1. Motivating Examples
   2. Modelling Strategy
   3. Results & Case Study
4. Discussion

Estimating SDG 4 Indicators with a Bayesian Modelling Framework
The education community has embraced the use of household surveys to supplement traditional data sources.

We have developed methods to harmonize survey data across different providers (e.g., MICS, DHS, national surveys) with administrative data in order to estimate:

1. SDG Indicator 4.1.2: Completion Rates
2. SDG Indicator 4.1.4: Out-of-School Rates

Our methods were developed with the following objectives in mind:

1. Consolidate data across sources into a single reliable, coherent series (point and interval estimates) for each country-indicator (including years without data).
2. Efficiently use scarce survey information.
3. Understand the relative reliability of different data sources in the education space.
Completion Rate Estimation

Key Ideas:

- Completion rates are estimated with a latent time series model, then adjusted for survey bias, late completion, age misreporting, and a differential error structure.
- Retrospective data reconstruction allows each survey to contribute ~20 years worth of information.

Out-of-School Rates

Definition: The “proportion of children and young people in the official age range for the given level of education who are not enrolled in pre-primary, primary, secondary or higher levels of education”

- Typically, administrative data provided by Ministries of Education are used to estimate enrollment quantities but these administrative data are not perfect and could benefit from supporting information.
  - Administrative data faces a two-source challenge due to separate population and enrollment sourcing.
  - Some countries have large gaps in administrative data or no data at all.
  - Household surveys are an attractive supporting data source to fill in the gaps though the infrequency is still a challenge.

- Objective: Consolidate these data sources and produce reliable OOS estimates for all countries with data in the years 2000-2020.
Motivating Examples

Estimating SDG 4 Indicators with a Bayesian Modelling Framework
Modelling Strategy

Key Ideas:

- Children and young people progress through school in cohorts, thus we build our model on age-specific cohort trajectories.
  - This specification is highly flexible and permits substantial variability in patterns across time and country.
  - Cohorting enables efficient data utilization - a survey may only cover one year, but it intersects ~12 cohorts.
- Administrative and survey data are fundamentally different with distinct challenges. Accordingly, we incorporate them into the estimation process through distinct mechanisms.

A first set of interactive results alongside an informal methods proposal have been released. A formal technical paper submission will follow in the near future.
At the global level, stagnation is the new story - the 2030 goal is in jeopardy.

In Sub-Saharan Africa, OOS rates are declining but the total number of OOS children is increasing.
Country-Level Results

Estimating SDG 4 Indicators with a Bayesian Modelling Framework
Case Study: DR Congo

Estimating SDG 4 Indicators with a Bayesian Modelling Framework
Case Study: DR Congo - By Year

Estimating SDG 4 Indicators with a Bayesian Modelling Framework
## Case Study: DR Congo - By Cohort

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>6</td>
<td>12</td>
<td>6</td>
<td>12</td>
<td>6</td>
<td>12</td>
<td>6</td>
<td>12</td>
<td>6</td>
<td>12</td>
<td>6</td>
<td>12</td>
<td>6</td>
<td>12</td>
<td>6</td>
<td>12</td>
<td>6</td>
<td>12</td>
<td>6</td>
<td>12</td>
<td>6</td>
<td>12</td>
<td>6</td>
<td>12</td>
<td>6</td>
</tr>
</tbody>
</table>

### Source
- Admin
- Census
- DHS
- MICS
- Other

---

Estimating SDG 4 Indicators with a Bayesian Modelling Framework
Active Projects

- An end-of-year update for both indicators is planned focusing on surveys released in the past year, as well as further introducing LIS and LFS data.
  - Results will be published on the Visualizing Indicators of Education for the World (VIEW) website.
- We are refining the presentation of results such that users can view observed and estimated values side-by-side more clearly.
- We are developing a more robust method of disaggregating by sex.
Closing

Future directions:

1. Education systems can be fragile and thus we should attempt to estimate their effects directly. However, given the heavy noise in this data, a crisis library is required to identify the country-years in crisis in advance.

2. Extend population disaggregations to other characteristics of interest and potentially begin to think about exploring subnational education estimation.

3. Build a complete short-term forecasting model for each indicator to facilitate discussion and planning.

4. Develop a unified education status framework that maps out-of-school rate estimates to completion rate estimates.

Thanks for listening! Any questions?