Institute for Statistics
Technical Cooperation Group

TCG Working Group on
Household Surveys

WG/HHS/2

OUT-OF-SCHOOL RATE MODELLING[1]

[1] This note was prepared by the Global Education Monitoring Report and UNESCO Institute for Statistics teams.

1. Context

Enrolment indicators are a traditional class of education indicators. Typically, administrative data is used to estimate various enrolment statistics, including out-of-school (OOS) rates, given their direct access to enrolment counts. However, administrative data is not without faults. Specifically, administrative data may be unaligned with other sources, incomplete, or conflict with population estimates (reported enrolments greater than population for example). In addition, administrative data are not available in many countries. As such, administrative data should be supplemented with additional data from independent sources when available.

This project seeks to supplant deterministic estimates of OOS rates produced using exclusively administrative data with probabilistic estimates that reconcile administrative data with household surveys and census data.

For the purposes of this project, an individual is defined as enrolled in school so long as they are enrolled at any grade level. They can be ahead of schedule or delayed by any number of years. This implies that OOS rate estimates will refer to individuals who are truly out-of-school. Ultimately, being delayed and being removed are two different problems that from a modelling perspective should be addressed separately. The quantification of delayed proportions will be estimated in a follow up to this project that seeks to perform a full accounting of the different categories of in-school students when the results of the OOS rate model are combined with completion rate estimates produced by the Adjusted Bayesian Completion (ABC) Rate model.

2. Data challenges

As discussed, the motivation behind this project is to seamlessly consolidate information from administrative systems with household surveys and censuses. The key data related challenges to be addressed are:

- Administrative data is provided by UIS whereas population data is sourced from WPP, resulting in frequent discrepancies. Specifically, 159 of 186 countries have negative OOS observations resulting from higher enrolments counts than populations. In total 20% of all observations are negative.
- Administrative data is not complete, it may be missing for certain periods, for certain ages, or entirely for certain countries.
- Administrative data may differ from survey data in level or trend.
- Survey data is limited by the infrequency of household surveys.
- Survey data accounts for approximately 15% of the data, the rest is administrative.

- Household surveys are occasionally highly biased and generally face greater uncertainty.

3. Modelling challenges

In addition to data challenges, the highly variable nature of OOS rates themselves introduces a number of key considerations, including:

- Flexibility is a top priority – we do not want to assume patterns for OOS rates across ages.
- Given that an individual can be considered in-school but enrolled in an earlier level than expected, dividing along education level is not appropriate. We choose to model ages directly.
- To enforce a strict [0,1] constraint on underlying true OOS rates, we estimate in a transformed space (specifically probit). However, administrative data cannot be transformed due to the presence of negative values resulting in a mismatch between administrative data and estimates.

4. Model

To estimate OOS rates probabilistically while also responding to the data concerns, we introduce a Bayesian hierarchical model that is thematically similar to the Adjusted Bayesian Completion (ABC) Rate model and to the Bayesian hierarchical models used frequently in health modelling but differs greatly due to the dynamic patterns characteristic of OOS rates. The model is divided into two stages. The first describes how underlying true OOS rates behave, and the second reconciles the true OOS rates with the observed data.

The first stage, the behaviour of underlying true OOS rates is composed of three drivers:

1. Baseline OOS rates: Initial enrolment drives OOS rates for all ages by setting the baseline from which late entry and dropout act upon. We choose a flexible time series model that seeks to capture the year-to-year variability as opposed to smoothing over it entirely.
2. Cohort dropout patterns: From the baseline OOS rates, we model changes in OOS that track students through their educational careers.

a. For example, for a cohort entering at age 6 in 2010, the OOS rate in 2010 is connected to the OOS rate in 2011 for age 7, then the OOS rate in 2012 for age 8, and so on.

b. We make no assumptions about the pattern of changes other than they must be smooth and transitions between patterns should also be gradual over time.

c. Tracing cohorts has positive consequences for forecasting. It is very natural to follow a cohort enrolled in school today along the expect cohort trajectory until completion.

3. Period shocks: Occasionally, short- to medium-term shocks to education systems do occur. We model these as rare impacts that decay in subsequent years at country-specific rates.

The second stage, the interaction of underlying OOS rates with observations, uses survey-specific bias terms, and source specific variability estimates. In addition, since survey data are non-negative, they can be transformed and interact with the estimated rate directly, promoting stability during estimation.

5. Discussion and next steps

The OOS model is capable of extracting patterns in OOS rates and producing coherent estimates for all countries with available data. The next steps are to:

- Finalize and validate the model (early 2022).
- Publish model results and forecasts with associated documentation and interactive visualizations for end users (early 2022).
- Integrate the OOS rate and completion rate estimates to produce detailed breakdowns of in-school populations resulting in a complete picture of education status (mid 2022).
- Address the root causes of the data challenges by collaborating with countries and education stakeholders to improve data reporting infrastructure and practices (second half of 2022).