Technical Cooperation Group
9th meeting
22-24 November 2022

WG/HHS/2

A Bayesian Cohort Model for Estimating SDG Indicator 4.1.4: OUT-OF-SCHOOL RATES

A proposal by the UNESCO Institute for Statistics and the Global Education Monitoring Report

TCG 9 Joint Working Group
on EMIS and Household Surveys

22 November 2022

A Bayesian Cohort Model for Estimating SDG Indicator 4.1.4: Out-of-School Rates

A proposal by the UNESCO Institute for Statistics and the Global Education Monitoring Report

August 28, 2022

Abstract

The out-of-school rate is an essential component of the Sustainable Development Goal (SDG) 4 monitoring framework used to track progress towards universal access to education. Producing estimates of out-of-school rates globally is a challenging task due to the diversity in enrolment patterns, systematic differences in the nature and reliability of administrative and survey-based data, and the heavy presence of invalid administrative observations resulting from enrollment counts that exceed corresponding population estimates. In this paper we introduce a cohort-based Bayesian hierarchical model to address these challenges and produce complete time series of out-of-school rates for 192 countries. The model uses a flexible spline-based process for underlying cohort out-of-school rate curves that are smoothed through cohort progression and over time. Observations are related to these values using a dual likelihood setup where each data source has distinct bias and variance components. The administrative side includes a structure that propagates uncertainty information contained in invalid data to avoid understating uncertainty. Validation exercises and sensitivity analysis suggest that the model is reasonably well calibrated and offers a material improvement over simpler approaches. The model is currently used by the United Nations to monitor out-of-school rates for all countries with available data.

Keywords: Bayesian modelling, household surveys, SDG 4, cohorts, administrative data

Contents

1 Introduction

As a key driver of social and economic mobility, promoting access to education is a core tenet of global development. Within the 2030 Agenda for Sustainable Development, the primary framework for global progress, the fourth of the 17 Sustainable Development Goals (SDGs) is concerned with "ensuring inclusive and equitable quality education and promoting lifelong learning opportunities for all" (United Nations 2015).

In pursuit of this objective, a global education monitoring apparatus consisting of a collection of indicators has been developed. These indicators are a window into the progress of nations on the path to universal high-quality education. Recently, two trends in the direction of education monitoring have emerged. The first is the modernization of classical education indicators using new statistical techniques and non-traditional data sources. The second is the development of new indicators focused on school completion and learning outcomes.

At the centre of the first trend is the family of enrolment indicators. Enrolment indicators, including the net enrolment rate, the gross enrolment ratio, and the out-of-school rate quantify the population of children or young people included or excluded from an education system.

Here we focus on the out-of-school rate. The UNESCO Institute for Statistics (UIS) defines SDG Indicator 4.1.4: Out-of-school rate as the "proportion of children and young people in the official age range for the given level of education who are not enroled in pre-primary, primary, secondary or higher levels of education"

(UNESCO Institute of Statistics 2022b). The importance of this indicator is quite clear - to achieve universal education by 2030 as specified by SDG 4, the out-of-school rate must be reduced to zero.

In recent decades, there has been considerable progress towards improving enrolment though the global community is far from universal enrolment. In 2018, UIS estimated that the total number of out-of-school children and youth globally still numbered approximately 258 million (UIS 2018). As the 2030 target draws near and educational expansion programs continue to be implemented, reliable estimates of out-of-school rates will be necessary to ensure timely progress and to guide policy interventions where needed.

Estimating these out-of-school rate for monitoring, however, is unfortunately fraught with challenges. Historically, estimates of out-of-school rates, and enrolment indicators more broadly, have relied exclusively on administrative enrolment counts produced by Ministries of Education around the world and corresponding population estimates. These administrative estimates, however, suffer from two key challenges. In many countries with high out-of-school rates, data reporting is incomplete or inaccurate, and in some country-years, there is no reporting at all. Second, given the two-source nature of the administrative estimates, there are often inconsistencies between enrolment counts and population estimates leading to negative out-of-school rate values. Given that negative values cannot be correct, these values are customarily dropped, set to zero, or capped before reporting. Yet this solution is not sufficient as it does not recognize that the root of the problem, that is the two-source nature of the data, also applies to positive out-of-school rate estimates.

In response to these challenges, there has been a recent push towards supplementing administrative data with censuses and other household surveys (Independent Expert Advisory Group on the Data Revolution for Sustainable Development 2014; UNESCO 2016). Survey-based data, however, cannot replace administrative data altogether as it tends to be infrequent and subject to survey bias, and sampling and non-sampling errors.

Spanning both types of data is the additional problem of occasional definition inconsistencies in out-of-school rate reporting and estimation. The official definition considers all youth enroled in school regardless of whether they are ahead, on-time, and behind schedule so as to differentiate between the problems of delayed education and removal from education, quantifying only the latter. Other definitions may exclude delayed students, thus bundling delays and removals into one figure.

Despite being a cornerstone of education monitoring, there has yet to be a standard approach to consolidate out-of-school rate estimates from the different data sources according to the official definition.

In this paper, we introduce a new Bayesian hierarchical cohort-based model that fills this need and estimates out-of-school rates for all countries with available data. The model estimates latent out-of-school rates using a spline-based cohort setup that mirrors the natural progression of students through a school cycle. Data from administrative and survey-based sources are then reconciled with a dual likelihood setup that recognizes the fundamental differences in the nature and data generation of the two classes of data while sharing information about bias and variance across countries. The model is now used by the United Nations to monitor SDG indicator 4.1.4 and track regional and global progress.

Similar styles of Bayesian hierarchical models have found success in estimating other development indicators. Within the education space, we estimated school completion rates in Dharamshi et al. (2022) using the Adjusted Bayesian Completion Rates (ABCR) model. The global health community has also found success applying such models to estimate under-5, neo-natal, and maternal mortality rates (Alkema and New 2014; Alexander and Alkema 2018; Alkema et al. 2016), stillbirth rates (Wang et al. 2020), and sex-ratios at birth

(Chao et al. 2021).

While the proposed model does have some structural similarities with prior indicator modelling work, primarily shared scaling terms, it introduces a few new key ideas. First, the latent process uses a cohorting system to link out-of-school rates together. This has some similarities with demographic modelling of population processes using cohort component models (Leslie 1945; UNPD 2019b). However, due to the lack of reliable data on the risks of late entry, dropout, repetition, and other relevant education transitions, we do not decompose changes in out-of-school rates and instead estimate the net grade-to-grade changes flexibly such that the underlying out-of-school rate cohort curves are smooth. Second, the dual likelihood setup is a new approach that accommodates the unique challenges of the two types of data. Specifically, the administrative data side is designed to accept negative observations as valid inputs while simultaneously constraining latent out-of-school rates to the $[0, 1]$ interval to allow information on data reliability to propagate from negative observations to positive observations. To our knowledge, this is the first attempt to contend with the invalid negative administrative data in a principled manner that avoids discarding information.

A final subject matter contribution, our proposed model shifts the focus from out-of-school rates for each of the three core education levels (primary, lower secondary, upper secondary), and estimates out-of-school rates by age. In practice, students enter and exit school at every age and thus there can be substantial variation within education levels. Consider the example of Burkina Faso in Figure 1. The "Primary" facet plots the observed out-of-school rates for all primary school children in each year. The remaining facets plot observed out-of-school rates for individual ages. The black line indicates a school policy change in 2010 that shifted the primary school age bracket from 7-12-year-olds to 6-11-year-olds. Notice that observed out-of-school rates decline significantly from ages 6 through 9 before rising again in subsequent ages. Our decision to estimate out-of-school rates at the individual age-level allows the model to capture such patterns while still permitting reporting by education level through post-processing aggregation steps.

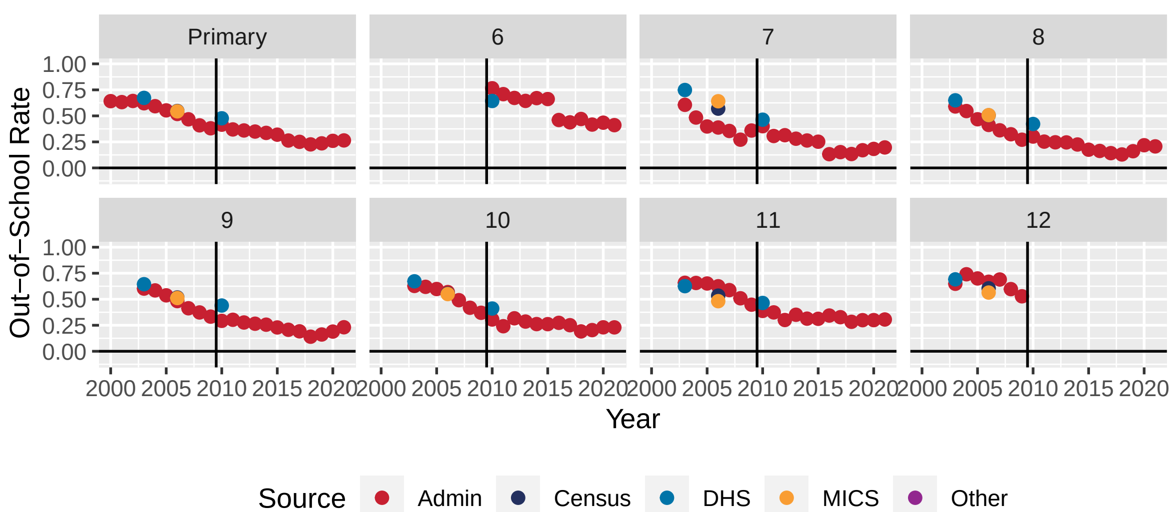

Figure 1: Obsevations of primary out-of-school rates for Burkina Faso. Each observation is coloured according to the corresponding source. A school policy change in 2010 adding 6-year-olds and removing 12-year-olds from primary school is denoted by the vertical black line.

The remainder of the paper is structured as follows. We begin by discussing the data and data challenges

Table 1: Observations of out-of-school rates by source type

Data Source	Observations	Percentage (%)
Administrative	92 036	83.5
Census	4 254	3.9
DHS	6 030	5.5
MICS	6 197	5.6
Other	1 689	1.5

in more detail in Section 2. The model specification is described in Section 3 with the underlying process model in Section 3.2 and the data model in Section 3.3. We then present results, validation exercises, and sensitivity analysis in Section 4 before concluding with a discussion in Section 5.

2 Data

Data on out-of-school rates can be classified into two broad categories: administrative and survey-based. Administrative data consists of official school enrolment counts by age and sex provided by Ministries of Education and compiled by the UNESCO Institute for Statistics (UIS). These enrolment counts are combined with corresponding population estimates produced by Eurostat for European countries (Eurostat 2000-2021), SingStat for Singapore (SingStat 2000-2021), and the World Population Prospects (WPP) (UNPD 2019a) for all other countries to compute out-of-school rates. Survey data include censuses, international survey programs, and other national household survey programs. Survey-based estimates of out-of-school rates, in contrast with administrative estimates, are single-source and are computed by aggregating responses to school attendance questions from the microdata. In total, there are 110 206 out-of-school rate observations across 192 countries with the breakdown of data by source provided in Table 1.

The dominance of administrative data in Table 1 is expected. Unlike household surveys which are conducted relatively infrequently, administrative data collection is intended to be continuous, providing annual updates to enrolment indicators. In practice, on average 121 countries provide administrative data updates annually compared with an average of 24 surveys per year in our dataset. This weighting between sources, however, is not consistent across all countries. Some countries have exclusively administrative data and some have exclusively survey data, but most contain some mix of the two. The observations from an illustrative set of countries are plotted in Figure 2.

Administrative data is clearly dominant in countries such as Argentina, Burkina Faso, Sweden, and Morocco but exhibits large gaps or is completely absent in the cases of Chad and the D.R. Congo. These latter examples reinforce the need to supplement the administrative data with survey-based data going forward.

2.1 Negative Observations

In addition to the lack of administrative data in certain countries, a second key challenge with administrative out-of-school data is a consequence of the two-source nature of the observations. Recall that administrative observations are generated by dividing administrative enrolment counts by separately source population estimates, then subtracting from one. As neither enrolment counts nor population estimates are error-free, it is possible, and indeed is unfortunately common, for population estimates in a given age-year to exceed

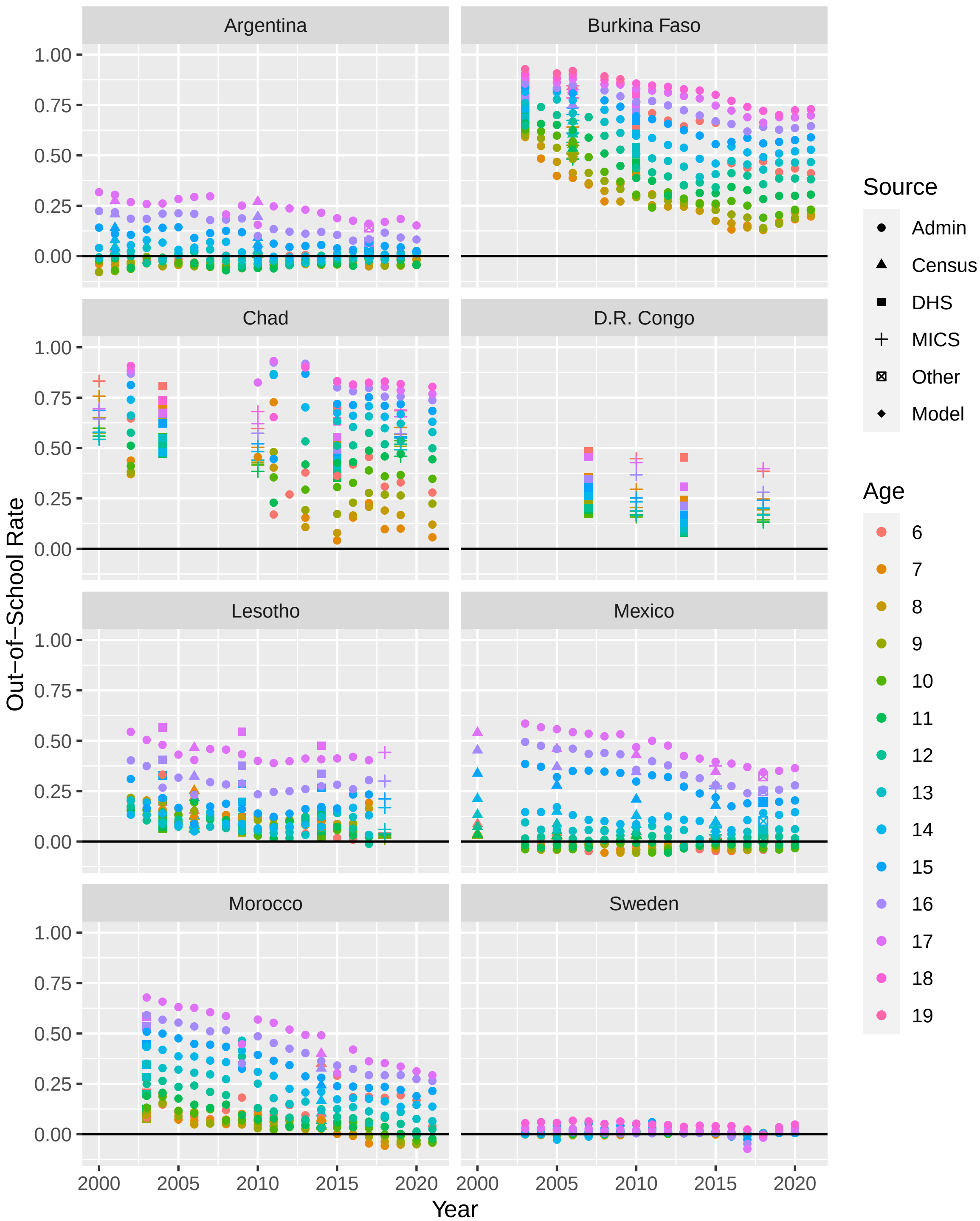

Figure 2: Observed out-of-school rates for a subset of countries. Each observation is presented by a distinct point. The age and source of each observation can be identified by colour and shape respectively.

the corresponding enrolment counts resulting in negative out-of-school rates. Examples of these negative observations are present in Argentina, Sweden, Mexico, and Morocco in Figure 2.

Negative observations are not exclusive to the above five countries. In total, 169 countries have at least one negative observation totaling 15.4% of the data. Further, negative observations do not tend to occur in isolation, but rather present as a consistent year-over-year, age-over-age phenomena as illustrated by the out-of-school rate observations for Argentina in Figure 3. This behaviour suggests that the cause of these observations is not isolated cases of age-misreporting or multi-year aggregation, but rather is a structural mismatch between the enrolment counts and population estimates.

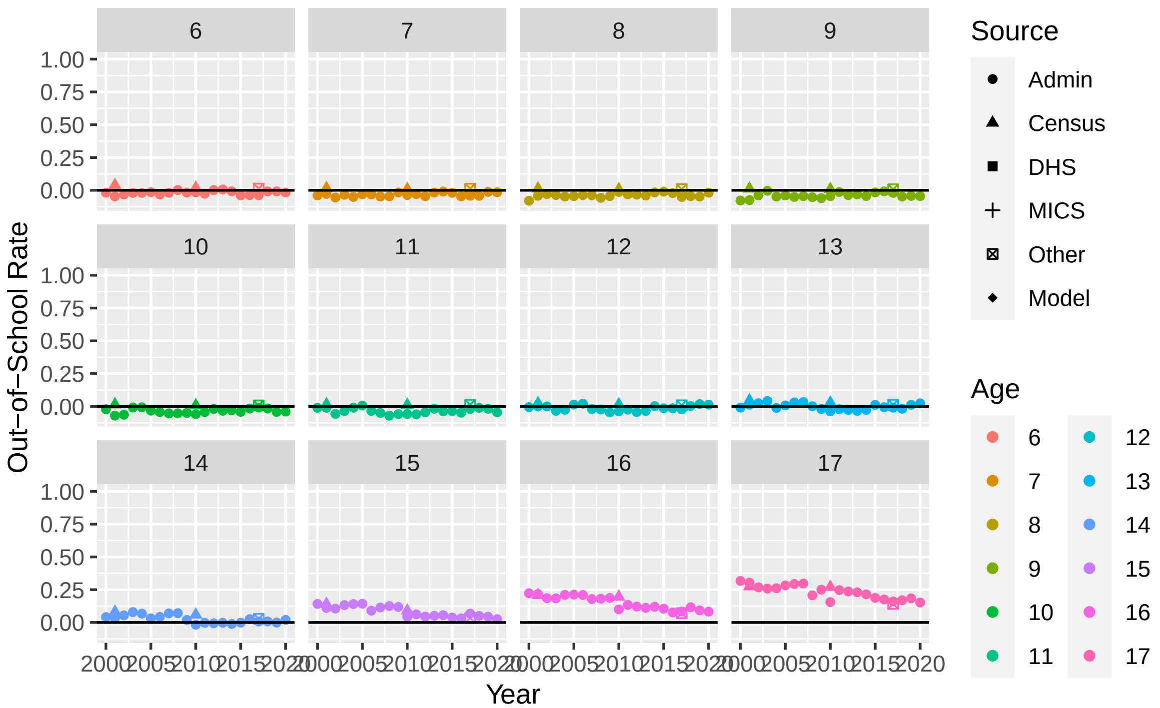

Figure 3: Observed out-of-school rates by age for Argentina. Each panel corresponds to a single age. The age and source of each observation can be identified by colour and shape respectively.

Customarily, negative observations are either dropped, replaced by zeroes, or capped as a data pre-processing step (UNESCO Institute of Statistics 2022a). The intuition behind these steps is that true out-of-school rates are confined to the $[0, 1]$ interval and thus adjusting the data to 0 could be a reasonable choice.

However, given that the negative administrative observations as a whole represent nearly as many observations as entire survey data class, and that they are generated in the same manner as positive administrative observations, we do not believe that a heavy-handed approach such as replacing or discarding all negative observations is advisable. Rather, we note that there is critical information in the negative observations. Specifically, in addition to suggesting that the true out-of-school rate values are low or approaching zero, they can offer insight into the error scaling of the entire administrative data class. As such, we have taken a more careful two-stage strategy. First, extreme observations defined as those under -0.1 were excluded. The remaining negative observations are left in the dataset to be addressed by the modelling strategy. Further discussion on data pre-processing and exclusion are provided in Appendix A.2.

3 Methods

The purpose of developing a model-based approach to estimating out-of-school rates is to consolidate administrative and survey observations, and produce coherent and complete estimates for all countries with data from the year 2000 onwards. Central to this objective is the reconciliation of two fundamentally different types of data. Out-of-school rates in reality cover virtually the entire $[0, 1]$ interval but are confined to that interval. As discussed in Section 2.1, administrative data does not abide by this constraint but survey data does.

To reflect these constraints, we will first distinguish between the two types of data by denoting A_i as an administrative observations for country $c[i]$, region $r[i]$, year $y[i]$, and age $a[i]$, then denoting S_j as a survey observation for country $c[j]$, year $y[j]$, age $a[j]$, survey wave $w[j]$, and survey series $s[j]$. Note that countries are nested within regions and survey waves are nested within survey series. Note that going forward, for simplicity, $a = 1$ corresponds to the youngest school entry age for a given country and higher values of a count up from the entry age. This is close to but not exactly equating age with grade. In a subset of countries, the entry age for grade one may change during the time interval, but to maintain consistency, we begin counting at the minimum entry age across all years and then aggregate the appropriate ages into education levels in keeping with changes in school schedules in post-processing.

We then, we define $\kappa_{c,y,a}$ as the target of estimation, the underlying true out-of-school rate for country c, year y, and age a, and constrain $\kappa_{c,y,a}$ by modelling $\Phi^{-1}(\kappa_{c,y,a})$ where Φ^{-1} is the probit function.

3.1 Model Summary

We begin by presenting an overview of the out-of-school rate model before describing the details of the process model and data model in Section 3.2 and Section 3.3 respectively.

Once again noting the fundamental differences in data generation and constraints, we consider a dual likelihood setup where administrative data are modelled in the real space and survey data are modelled in the probit space as follows:

$$A_i \sim \mathcal{N}\left(\kappa_{c[i],y[i],a[i]}, \sigma^2 + \omega^2_{r[i]}\right), \tag{1}$$

$$\Phi^{-1}(S_j) \sim \mathcal{N}\left(\Phi^{-1}\left(\kappa_{c[j],y[j],a[j]}\right) + \beta_{c[j],s[j]}, \frac{\nu_j^2 + \eta^2_{w[j]}}{f\left(\Phi^{-1}(S_j)\right)^2}\right), \tag{2}$$

where $\kappa_{c,y,a}$ is the underlying true out-of-school rate for country c, year y, and age a, σ is the global administrative data error, ω_r is the regional administrative data error, $\beta_{c,s}$ is the survey bias for country c's surveys originating from survey series s, ν_j is the sampling error for survey observation j, and η_w is the non-sampling variance for survey wave w.

The intuition behind Equations (1) and (2) is as follows. For administrative data, since $\kappa_{c,y,a}$ is lower-bounded at 0, the likelihood of negative observations can only be increased by pulling $\kappa_{c,y,a}$ values towards zero or by expanding the error terms σ and ω_r. As the error terms are shared, expanded variances propagates to the rest of the administrative data, calibrating the uncertainty in a way that would not be possible if such data points were excluded. Equation (2) is entirely in the probit space, reflecting the intrinsic $[0, 1]$ constraint on survey-based estimates. Survey-based estimates of education indicators, however, are expected to suffer from some amount of survey bias and are subject to sampling and non-sampling variances (Dharamshi et al. 2022).

3.2 Process Model

Mirroring the natural progression of students through a school system, we model $\kappa_{c,y,a}$ using a cohorting structure. Cohorting is a common technique used in demographic modelling, often tracing birth cohorts over time until death (UNPD 2019b). The education context differs in that not all students enter school at the entry age and thus the population out-of-school can, and often does, decrease before dropout begins to trigger increases. Furthermore, such patterns change over time. As countries enact policies that target elevated dropout rates or depressed early childhood enrolment, the out-of-school rate cohort curves shift in response. As such, the specification for κ must be flexible, but also gradual, along both the age and time dimensions. A primary consequence of selecting a cohort-based approach is that observations primarily thought about along a cohort dimension, not along the time dimension. Visually, on the age-time plots of Figure 2, observations are considered 'diagonally', not 'vertically'.

To capture these dynamics, we propose a two-component setup to model $\kappa_{c,y,a}$. The first tracks entry age out-of-school rates over time and the second captures changes in out-of-school rates along cohort progression trajectories.

The summary equation governing the $\kappa_{c,y,a}$ process is as follows:

$$\Phi^{-1}\left(\kappa_{c,y,a}\right) = \mu_{c,y-a+1} + \Theta_{c,y,a}, \tag{3}$$

where $\mu_{c,y-a+1}$ is the entry age out-of-school rate for country c in the year $y-a+1$, and $\Theta_{c,y,a}$ is the accumulation of changes in out-of-school rate for the cohort that entered into school in year $y-a+1$ progressed through the school system and reached age a by the time of observation in year y.

3.2.1 Baseline Out-of-School

We model the first term, $\mu_{c,y}$, as a random walk with drift process smoothed using splines with annual knots. The smoothed random walk with drift setup allows for both long-term trends in entry-age enrolment as well as shorter-term shocks representing annual volatility in enrolments. The model equations and prior distributions are as follows:

$$\mu_{c,y} = \sum_{k=1}^{K_1}\left(B_k^{(1)}(y)\mu^*_{c,k}\right), \tag{4}$$

$$\mu^*_{c,k+1} = \gamma_c + \mu^*_{c,k} + \epsilon_{c,k}, \tag{5}$$

$$\gamma_c \sim \mathcal{N}^-\left(0, \sigma^2_{\gamma,r[c]}\right), \tag{6}$$

$$\sigma_{\gamma,r[c]} \sim \mathcal{N}^+\left(0, 0.02^2\right), \tag{7}$$

$$\epsilon_{c,k} \sim \mathcal{N}\left(0, \sigma^2_\mu\right), \tag{8}$$

$$\sigma_\mu \sim \mathcal{N}^+\left(0, 0.1^2\right). \tag{9}$$

The long-term drift, γ_c is assumed to be negative, implying progress towards universal education on the enrolment front. The scaling of the trend is assumed to regional allowing for similar countries to exhibit similar behaviour. For example, highly developed regions with universal primary enrolment will share a near-zero scaling on γ_c. The specification of a spline setup with annual knots for smoothing was selected to balance between the flexibility of the random walk with the risk of $\epsilon_{c,k}$ overfitting potentially erratic data. This setup allows $\mu_{c,y}$ to be responsive to shocks so long as there is some evidence of short-term persistence. We study the consequences of this modelling decision by testing the model against versions with simpler specifications for $\mu_{c,y}$ in Section 4.4.

3.2.2 Cohorting

In contrast to the entry-age out-of-school rates, we do not model higher-age out-of-school rates directly, rather the model is structured around the changes in out-of-school rates as a cohort ages. A specification built around first differences in out-of-school rates better reflects our understanding of the behaviour of youth in school systems. Intuitively, if a student drops out of school today, our baseline expectation is that they will remain removed from school system tomorrow. That is, an increase in out-of-school rate should propagate along the cohort out-of-school curve as the cohort progresses forward. Similarly, if a student enters school late triggering a decrease in the out-of-school rate, that too should propagate as the cohort progresses. These behaviours are captured using a vector-valued P-splines setup as follows:

$$\Theta_{c,y,a} = \sum_{g=1}^{a-1} \sum_{k=1}^{K_5} \left(B_k^{(5)}(y-a+1)\alpha_{c,g,k} \right), \tag{10}$$

where $\alpha_{c,g,k}$ corresponds to the change in out-of-school rate for country c, age g, and spline k, and $B_k^{(5)}$ are the spline coefficients corresponding to a spline setup with 5-year intervals between knots. The specification for $\Theta_{c,y,a}$ in Equation (10) can be understood in two pieces defined by the summations. The inner summation is the standard spline-based smoother that induces gradual movement between the dominant out-of-school rate first order difference patterns every 5-years. We note that the spline coefficients correspond to, $y - a + 1$, the entry-year for the cohort, thus inducing the cohort structure. The choice of 5-years is intended to reflect the reality that policy implementation is not instantaneous, but can take many years. Given the central nature of this assumption, we also conduct sensitivity analysis on the number of years between knots in Section 4.5.

Following this step, the outer summation then aggregates the changes along the cohort curve to produce the net change in out-of-school rate from the entry-age out-of-school rates defined by $\mu_{c,y}$.

Viewed as a whole, the $\alpha_{c,g,k}$ parameters form an $(\max(a) - 1) \times K_5$ matrix where each column characterizes one set of cohort out-of-school rate changes and each row contains the set of all changes between two specific ages over time. We then smooth along both the columns and rows of this matrix as follows:

$$\alpha_{c,g,k} \sim \mathcal{N}\left(\alpha_{c,g-1,k}, \tau_{\alpha_1}^{-1}\right), \tag{11}$$

$$\alpha_{c,g,k} \sim \mathcal{N}\left(\alpha_{c,g,k-1}, \tau_{\alpha_2}^{-1}\right), \tag{12}$$

$$\sum_{g=1}^{a-1} \alpha_{c,g,k} \sim \mathcal{N}\left(0, 2^2\right), \tag{13}$$

$$\tau_{\alpha_1} \sim \mathcal{N}^+\left(0, 1^2\right), \tag{14}$$

$$\tau_{\alpha_2} \sim \mathcal{N}^+\left(0, 1^2\right), \tag{15}$$

where τ_{α_1} and τ_{α_2} are the column and row smoothing parameters equipped with vague priors. Along each column, τ_{α_1} smooths first differences such that the cohort out-of-school rate curves are effectively assigned RW(2) priors. Along each row, τ_{α_2} introduces a temporal smoothing effect. Given the grid-like structure and smoothing between neighbouring cells, this setup does have some structural similarities with a Conditional Autoregressive (CAR) model (Besag 1974). However, we note that in the out-of-school case, there is no spatial context, and we accumulate along columns to produce cohort curves. The accumulated column sums are then constrained such that all else equal, smaller net changes in out-of-school rate for each cohort are preferred.

We illustrate the mechanisms of this setup by plotting heatmaps of the $\alpha_{c,g,k}$ matrix for a illustrative selection of countries in Figure 4. Blue cells correspond to decreases in out-of-school rates, red cells correspond to increases in out-of-school rates, and white cells represent stable out-of-school rates. Note that these plots exclusively plot $\alpha_{c,g,k}$. They are not anchored by $\mu_{c,y}$ values and thus do not represent out-of-school rate *values*, they only provide a window into out-of-school rate *changes*.

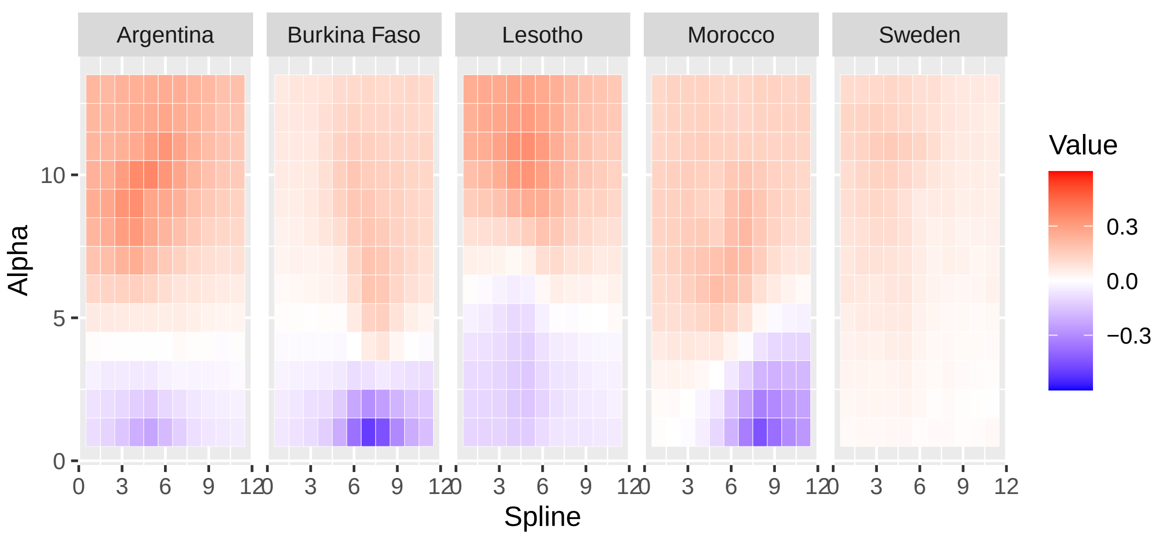

Figure 4: Change in out-of-school rate heatmaps for Argentina, Burkina Faso, Lesotho, Morocco, and Sweden.

The examples in Figure 4 illustrate the flexibility of the model to capture a diverse set of patterns. The plot for Argentina illustrates a relatively stable pattern over time with a net decrease in out-of-school at the youngest ages followed by a minimum in out-of-school in the fourth alpha row and concluding with a steady increase in out-of-school in the older ages. Burkina Faso and Lesotho present dramatically different patterns where there are many negative changes in out-of-school rates before peak cohort enrolment is achieved. Morocco, by contrast, illustrates a case with substantial changes in patterns. Over time, the trough in out-of-school rates indicated by the white cells is shifting to older ages suggesting growth in late entry. Finally, the $\alpha_{c,g,k}$ values for Sweden indicate no evidence of net late entry (ie. there are no blue cells) and overall there is minimal activity as expected in a mature education system.

3.3 Data Model

To relate the observed data to the underlying out-of-school rate process model, there are additional data effects that must be considered. Specifically, this model addresses source-specific differences in systematic bias and variance.

3.3.1 Bias

With respect to bias, we first assume that administrative data is not subject to a source-specific bias term. In contrast, school completion rates, a related education indicator also produced by household surveys, were found to suffer from material levels of structural bias (Dharamshi et al. 2022). This suggests that survey bias must be considered in modelling out-of-school rates. We thus include a country- and source-specific bias term, $\beta_{c,s}$, for survey data in Equation (2). $\beta_{c,s}$ is modeled hierarchically as $\beta_{c,s} \sim \mathcal{N}\left(0, \sigma^2_{\beta,s}\right)$ with $\sigma_{\beta,s} \sim \mathcal{N}^+\left(0, 0.1^2\right)$ such that source-specific scaling is shared across countries.

3.3.2 Variance

The variance structures assumed in Equations (1) and (2) are selected to reflect the specific characteristics of each source of data.

The administrative data variance breaks down into global and regional terms. The global component, σ, is modeled as $\sigma \sim \mathcal{N}^+\left(0, \sigma^2_\sigma\right)$ where $\sigma_\sigma \sim \mathcal{N}^+\left(0, 0.01^2\right)$. The regional piece, ω_r, is then modeled similarly as $\omega_r \sim \mathcal{N}^+\left(0, \sigma^2_\sigma\right)$. The inclusion of a regional term for administrative data error is intended to capture regional similarities in data infrastructure and developmental status. As an example, population data for all European countries is sourced from Eurostat suggesting that errors should be shared at this level. The global term then sets a floor on error levels for the data class as a whole. No matter how robust an individual country or region's data reporting system may be, all administrative observations of out-of-school rates suffer from the two-source problem and thus some level of mismatch is expected. The global term allows the error information contained in implausible negative observations to flow to the rest of the data, thus introducing a baseline level of skepticism.

For survey data, we divide the error in sampling and non-sampling components. Sampling errors, ν_j capture observation specific variation and are estimated from the microdata prior to modelling using the clustered Jackknife procedure used by DHS to estimate standard errors (The Demographic and Health Surveys Program 2012). These estimates are then supplied as inputs to the model. Non-sampling errors, η_w is then estimated within the modelling framework using a hierarchical structure. Specifically, $\eta_w \sim \mathcal{N}^+\left(0, \sigma^2_\sigma\right)$. In order to enhance comparability between data sources, both ν_j and η_w are estimated in the real space and then transformed to the probit space using a Delta method transformation. We illustrate the benefits of this setup by directly comparing posterior variance estimates in Section 4.3.

3.4 Implementation

Posterior samples of all model parameters are generated using the No-U-turn sampling (NUTS) Hamilton Monte Carlo algorithm (Hoffman and Gelman 2014; Neal 2011) implemented in the Stan R package (Carpenter et al. 2017; Stan Development Team 2020b). Four parallel changes are executed for each of the male, female, and total populations, each with 4 000 iterations total. The first 2 000 iterations of each chain are removed as burn-in. The remaining iterations are thinned due to memory constraints to a final total of 1 000 iterations per population. Convergence is assessed using the standard diagnostic checks including trace plots, effective sample size (ESS), and the Gelman and Rubin diagnostic (Gelman and Rubin 1992; Vehtari et al. 2020; Stan Development Team 2020a).

3.5 Validation

The performance of the model is assessed with out-of-sample validation exercises. We randomly drop 20% of the observations from each country and fit the model on the reduced dataset. We then compare the left-out observations to the corresponding median $\kappa_{c,y,a}$ values, adjusted for estimated survey bias where appropriate, and compute mean squared errors (MSE) and mean absolute errors (MAE). Given that negative observations implicitly impose a floor on error values as $\kappa_{c,y,a}$ is bounded below at zero, we also compute adjusted mean squared errors (aMSE) and adjusted mean absolute errors (aMAE) where errors on administrative data are computed as the difference between estimated values with $\max(0, A_i)$. Finally, we examine coverage of the prediction intervals defined as $n^{-1}\sum_{i=1}^{n}\mathbb{1}_{l_i \leq y_i \leq u_i}$ where n is the number of observations left-out, i is the current observation, and l_i and u_i are the lower and upper bounds of the prediction interval.

4 Results

4.1 Illustrative Examples

The results of the model for a representative set of countries is presented in Figure 5. For each country, there is one observation plot and one results plot with the complete set of model OOS rate estimates from 2000-2022. Each country is selected to highlight a distinct scenario with respect to data availability and patterns or trends in out-of-school rates. As is typical in education settings, there is great diversity across countries and thus the model must be flexible in response.

Considering the selection of countries from simplest to model to most complex, we begin with the example of Sweden. As a country with a highly developed education system, universal enrolment is observed and modeled with little uncertainty. Argentina is a slightly more complex case as there is non-insignificant out-of-school at the highest ages. However is relatively stable with small declines in recent years. Lesotho, Mexico, and Morocco are three similar examples with stable or declining year-over-year out-of-school rates over the time interval and near-zero out-of-school in younger ages in recent years. Morocco, however, is found to have stagnant entry age out-of-school rates as seen by the orange line in the "Morocco Estimated" facet. This is consistent with the growth in net negative changes found in the Morocco heatmaps in Figure 4.

Burkina Faso, Chad, and D.R. Congo are more complex examples with varying degrees of data completeness, agreement between sources, and complexity in out-of-school rate changes. With Burkina Faso, there is nearly complete administrative data reporting and consistency between administrative and survey-based data. However, while the youngest and oldest age out-of-school rates only experience moderate declines, there

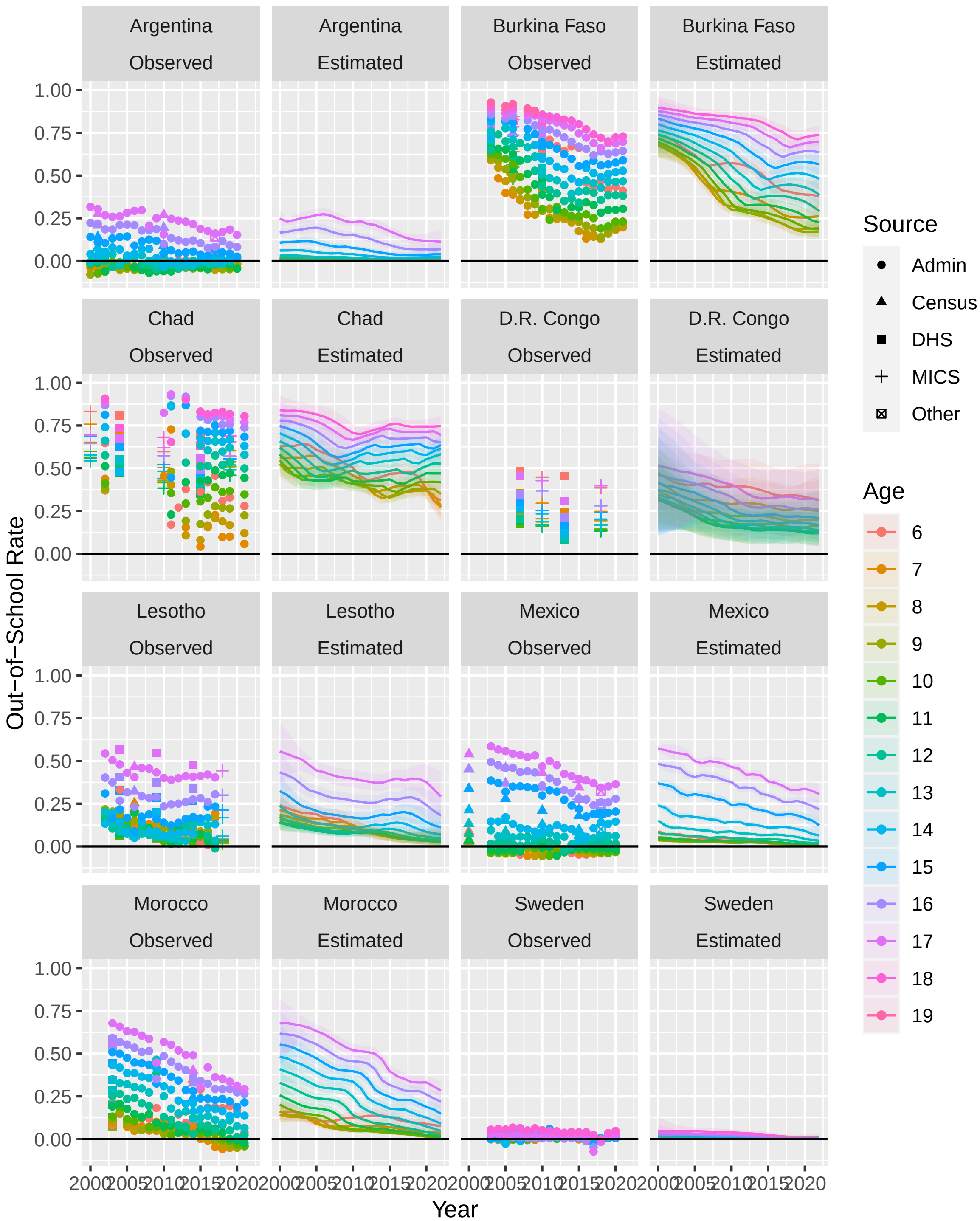

Figure 5: Observations and estimates of out-of-school rates for eight countries. Each country is assigned two panels. The first plots the observed out-of-school rates. The age and source of each data point can be identified with the legend to the right of the plot. The second panel plots the estimated out-of-school rate curves for each age along with the associated 90% uncertainty intervals. The age of each curve can be identified by the colour of the line and associated shaded region.

is substantial progress in improving enrolment in the middle ages. The flexibility model specification here allows for adequate pattern extraction in this case. Chad is a more difficult case where there is substantial disagreement between surveys and administrative data as well as a data gap between 2005 and 2010. In this case where there is large quantities of consistent survey-based data along with highly variable administrative data, the level of out-of-school reported by the surveys is generally preferred. However, we note that the uncertainty intervals in this case do expand to reflect the greater disagreement between data sources. Finally, in the case of D.R. Congo, there is no available administrative data and only four surveys, two DHS and two MICS. In this case, there is general alignment between the survey sources and as such, coherent estimates of out-of-school rates are produced by the model. However, due to the dearth of data, large uncertainty intervals are produced, particularly at the entry and exit ages.

The plots in Figure 5 offer a complete look at all of the estimates and observations at once. However, given the sheer number of points and curves, it can be difficult to identify underlying patterns within each country. In response to this challenge, in Figure 6, we focus on four specific cohorts to illustrate the diversity of patterns produced by the model for Argentina, Burkina Faso, Lesotho, Morocco, and Sweden. In each column, we plot the observations and estimates of out-of-school rates for the four cohorts that entered school in 2000, 2005, 2010, and 2015. The example of Argentina shows that while the pattern of minor increases in out-of-school at older age persists, there is overall a downwards shift over time. With Burkina Faso and to a lesser extent Lesotho, the distinct U-shape pattern suggested by the heatmaps in Figure 4 is quite clear. With Morocco, the entry age out-of-school rate remains stable over time but as indicated by the heatmaps, in later cohorts, progress in enroling missing students at later ages results in lower minimum out-of-school rates. Finally, the consistent near-zero estimates in the Sweden plots are indicative of a mature education system.

4.2 Aggregated Results

The low level country-year-age basis for the model allows for further transformations and aggregations of substantive interest. We illustrate this consequence with a global example. In Figure 7, estimates of out-of-school rates are converted to estimates of out-of-school population numbers by multiplying by the corresponding country-year-age population estimates. These values are then aggregated into each country and then to SDG regions to produce estimates of the total number of out-of-school children and youth by region. These values are plotted on a stacked area graph to illustrate the changing composition of the total out-of-school population globally.

The results of Figure 7 are both uplifting and disheartening. On the positive side, there has clearly been a substantial reduction in the global number of out-of-school children and youth from approximately 400 million to 250 million over the past two decades. This reduction has largely occurred in Central and Southern Asia, and Eastern and South-eastern Asia. While the decrease in undeniable, the *rate* of decline in the out-of-school population has slowed considerably in the past decade suggesting efforts to reduce exclusion from education are struggling. Further, in Sub-Saharan Africa, the number of out-of-school children is actually *increasing*, a finding of great concern for the global education community.

4.3 Posterior Variance Estimates

In addition to estimating out-of-school rates themselves, we noted in the model variance specification an interest in comparing error estimates across sources. In Figure 8, boxplots of the posterior estimates of total

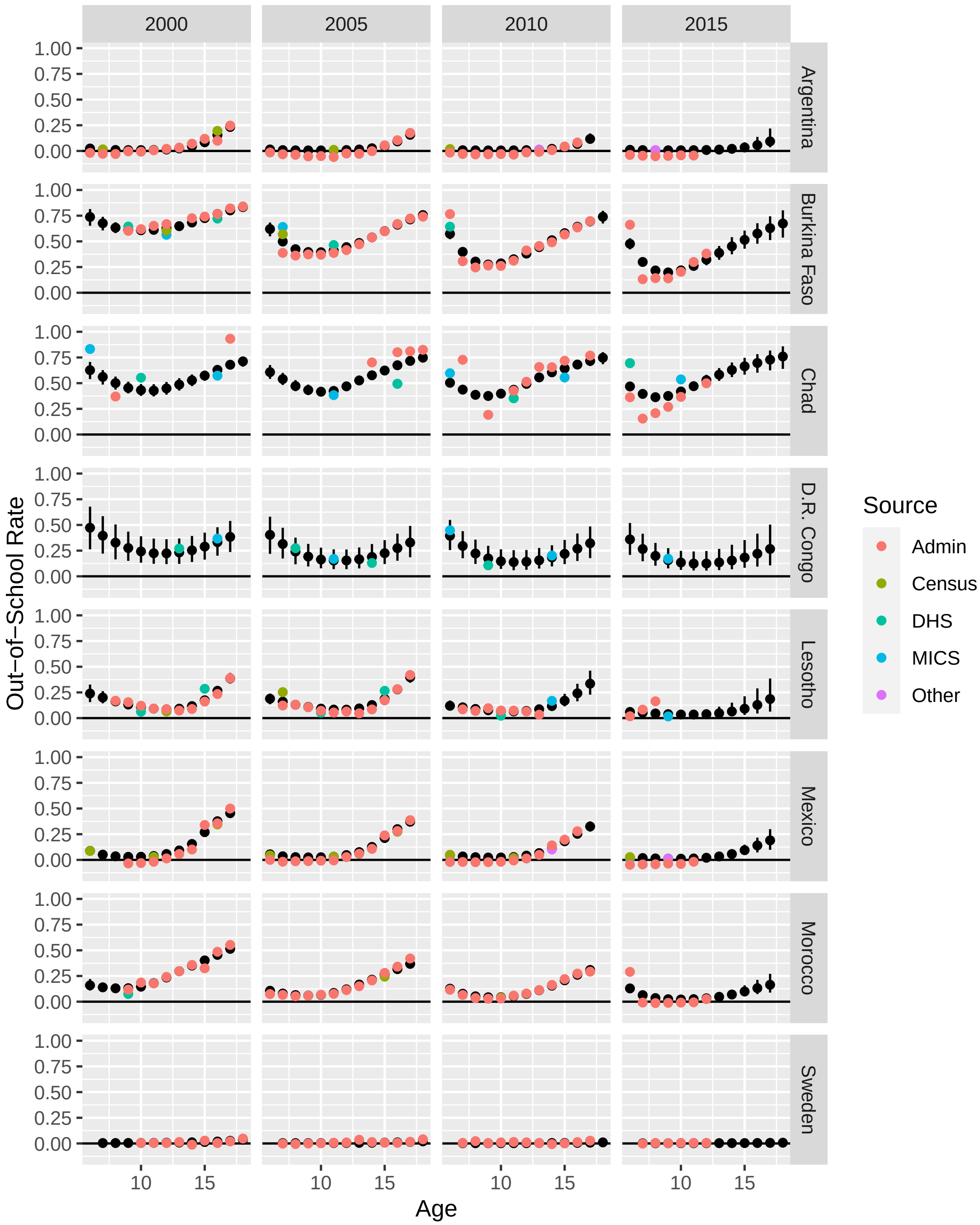

Figure 6: Estimates and observations of out-of-school rates for eight countries. Model-based estimates are presented by the black dots and the associated 90% uncertainty intervals are given by the vertical black lines. Observations are presented by coloured points and the source can be identified by the colours in the legend.

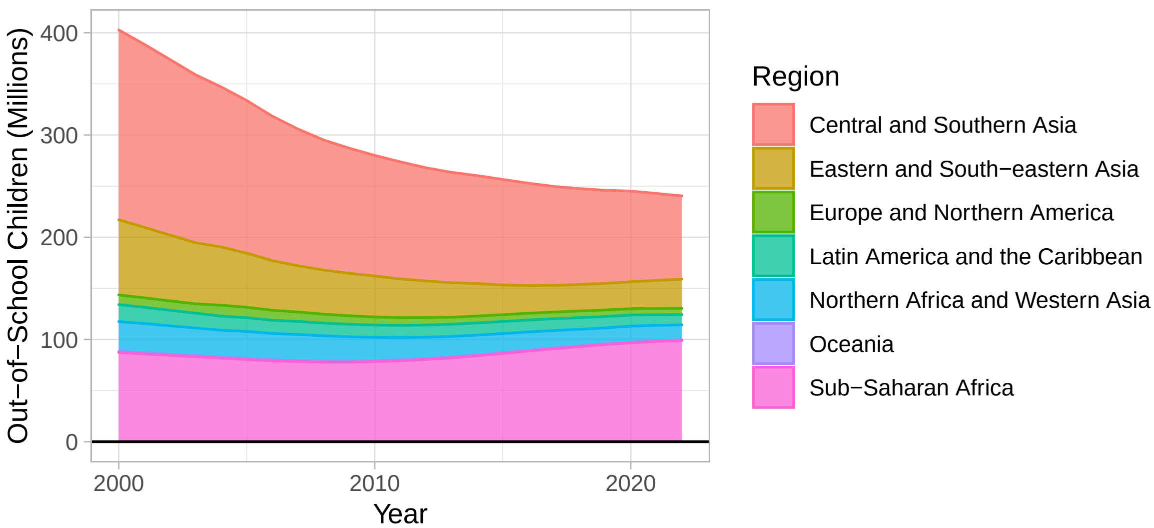

Figure 7: Out-of-school population by region.

standard error by source type are presented. In the case of administrative data, this is the sum $\sigma + \omega_r$ and for survey data, we plot $\nu_i + \eta_w$. Consistent with the survey data non-sampling variance specification, each wave of DHS and MICS surveys is plotted separately. Censuses and other surveys, however, are aggregated due to the relatively smaller number of observations per wave in these categories.

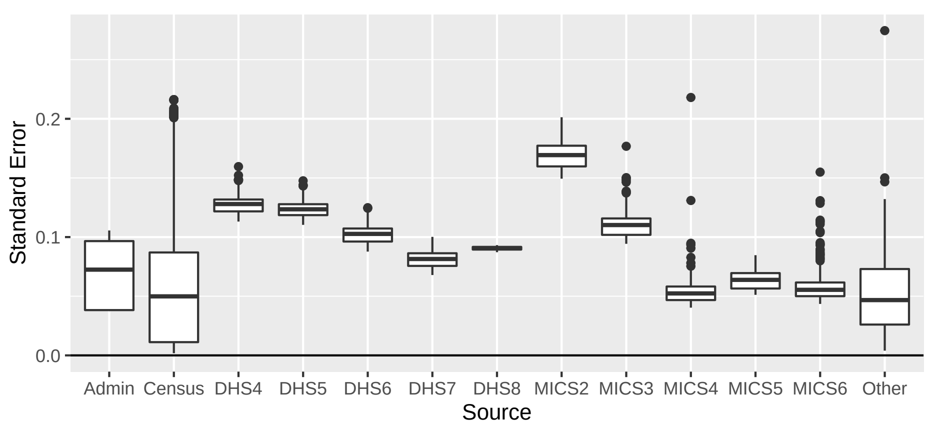

Figure 8: Distribution of model variance posterior medians by source.

The boxplots in Figure 8 tell a story about the relative reliability of out-of-school rate data sources. Over time, successive waves of DHS and MICS surveys have decreasing total standard errors. In the most recent MICS waves, the median standard error falls below that of administrative data, validating the broad push towards household surveys. That said, the distributions of survey data errors tend to be skewed with occasional extreme variance values corresponding to surveys presenting out-of-school rate observations that dramatically depart from the consensus of other sources either in magnitude or pattern. In the best cases, however, certain censuses and nationally run surveys are modeled as having extremely low error values. Focusing on magnitudes, it is worth noting that the median total standard error across all observations is 0.0725. The large error values reinforce the need to consider multiple data sources simultaneously in order to have comfort

Table 2: Out-of-sample validation results.

Model Version	MSE	aMSE	MAE	aMAE	Coverage (80%)	Coverage (90%)	Coverage (95%)
Model	0.0051	0.0046	0.0435	0.0389	0.88	0.93	0.95
MuK5	0.0050	0.0045	0.0439	0.0393	0.87	0.93	0.95
MuRW	0.0051	0.0046	0.0439	0.0393	0.88	0.93	0.95
Simple	0.0072	0.0069	0.0533	0.0487	0.75	0.80	0.82
NoBias	0.0057	0.0052	0.0459	0.0413	0.88	0.93	0.95

over resulting out-of-school estimates.

4.4 Validation

At present, there is no clear standard approach to estimating out-of-school rates to compare our model's results against. Direct estimates from the various data sources have historically been dominant in out-of-school monitoring use cases. This is a recurrent challenge in education modelling projects. However, as we noted in Dharamshi et al. (2022), simple linear models are applied in practice for SDG 4 indicator forecasting (EFA Global Monitoring Report 2015; UNESCO Institute for Statistics, Global Education Monitoring Report Team 2019). As such, we consider a 'Simple' Bayesian model where the out-of-school rate is modelled linearly in the probit space as a function of year and age. This approach contrasts sharply with our proposed model in that observations are linked 'vertically' within years as opposed to 'diagonally' through cohorts. Additionally, this approach does not consider a dual likelihood structure, but rather models administrative data alongside survey data in the probit space by setting negative values to 0.0032% (-4 in the probit space). This is consistent with the common practice of imputing zeros in place of negative observations. In addition to the basic model, we consider reduced or simplified versions of our proposed model to study the consequences of specific modelling decisions. In the 'NoBias' model, we drop the survey bias term $\beta_{c,s}$. In the 'MuRW' model, we replace the specification for $\mu_{c,y}$ with an unsmoothed random walk without drift. Finally, in the 'MuK5' model, we increasing the smoothing on the baseline out-of-school rate by increasing the interval between $\mu^*_{c,k}$ knots to 5 years such that it matches the smoothing on $\Theta_{c,y,a}$. The specifications of 'MuRW' and 'MuK5' are intended to create scenarios of reduced and increased smoothing on the baseline out-of-school rates respectively. The effects of assumptions regarding smoothing on the changes in out-of-school rates are discussed in Section 4.5.

Mean squared error (MSE) and mean absolute error (MAE) results for out-of-sample data from the validation exercises are presented in Table 2 along with coverage results for the entire sample.

Starting with the error metrics, the specified model outperforms on MAE and is in-line or better on MSE. The outperformance is most stark when considering the simple model. It is clear that the 'Simple' model has materially higher errors than the model suggesting that the introduced cohorting, likelihood, bias, and variance ideas are valuable in estimating out-of-school rates. To a lesser extent, similar results are observed in the 'NoBias' model. Once again, errors are higher than the complete model though given that the bias terms only affect a subset of observations, the magnitude of outperformance is smaller. The 'MuRW' and 'MuK5' results are interesting. The error performance only marginally departs from that of the model as specified in Section 3 suggesting that the specification of the underlying baseline out-of-school rate is not highly sensitive to the exact specification but is rather capturing the underlying patterns in the data.

Table 3: Sensitivity results.

Model Version	MSE	aMSE	MAE	aMAE	Coverage (80%)	Coverage (90%)	Coverage (95%)
Model	0.0051	0.0046	0.0435	0.0389	0.88	0.93	0.95
Tau1 (0.5 sd)	0.0049	0.0045	0.0428	0.0382	0.88	0.93	0.95
Tau1 (2 sd)	0.0052	0.0047	0.0444	0.0398	0.88	0.93	0.95
Tau2 (0.5 sd)	0.0050	0.0045	0.0430	0.0384	0.88	0.93	0.95
Tau2 (2 sd)	0.0051	0.0046	0.0441	0.0395	0.88	0.93	0.95
3-year knots	0.0052	0.0047	0.0442	0.0395	0.88	0.93	0.95
8-year knots	0.0050	0.0045	0.0434	0.0388	0.88	0.93	0.95

The coverage results in Figure 2 tell a similar story. The 'Simple' model has poor coverage further confirming its inadequacy in this context. The rest of the models, however, have coverage values generally in line with the nominal figures. The coverage of 80% and 90% prediction intervals is higher than expected suggesting a slight degree of overdispersion. Overall, the error and coverage results suggest that the proposed model specification offers a material improvement over simpler approaches to estimating out-of-school rates and is resonably well calibrated.

4.5 Sensitivity

In addition to the above validation exercises, we also test the sensitivity of the model to key assumptions and prior distributions, focusing on $\Theta_{c,y,a}$, the matrix responsible for capturing the changes in underlying true out-of-school rates. Specifically, we change the number of years between knot points from 5 years to 3 and 8 years to test the performance of the model under different levels of out-of-school rate pattern flexibility. We also consider sensitivity to the prior distributions assigned to τ_{α_1} and τ_{α_2}, the $\alpha_{c,g,k}$ smoothing terms, from $\mathcal{N}\left(0, 1^2\right)$ to $\mathcal{N}\left(0, 0.5^2\right)$ and $\mathcal{N}\left(0, 2^2\right)$. We then perform out-of-sample exercises by randomly leaving out 20% of the total population data from each country and fit each model variation. The results presented in Table 3 include out-of-sample MSE and MAE values and full sample coverage values.

The results of the sensitivity analysis suggest that the model is not highly sensitive to the specific prior distributions. Coverage values are consistent across all cases and error performance is comparable across specifications. There is a minor level of outperformance when the constraints on $\alpha_{c,g,k}$ are relaxed by lowering the smoothing priors to $\mathcal{N}\left(0, 0.5^2\right)$. This is not unexpected as relaxed smoothing constraints allow the model to accommodate more noise in the data. The interval between knot points paints a more complex picture. One might expect that with more knots, the model would be more flexible, and thus perform better. However, we find that the 3 year interval slightly underperforms the 5 and 8 year intervals suggesting that shrinking the interval would lead to overfitting. These results validate the assertion that a 5 year interval adequately balances flexibility and performance while remaining plausible from a policy implementation cycle perspective.

5 Discussion

In this paper we propose a new method to estimate SDG global indicator 4.1.4: out-of-school rates. Our model harmonises administrative and household survey-based data within a cohort-based Bayesian hierarchical framework to produce complete estimates for 192 countries from 2000 to 2022. The proposed specification is compared against a variety of simpler models and also tested for sensitivity to key assumptions. The results from these exercises validate the specification and indicate that the model is reasonably well calibrated. The United Nations has now adopted our model to produce estimates of SDG indicator 4.1.4 and to monitor country, regional, and global progress.

Estimating out-of-school rates is a challenging task due to the diversity in plausible enrolment patterns, systematic differences between administrative and survey-based data, and the presence of negative administrative out-of-school rate observations caused by enrollment counts that exceed corresponding population estimates. We designed our approach to address each of these key challenges. The diversity of out-of-school rate patterns inherent in the education context is captured using a flexible vector-valued spline-based setup that models the changes in out-of-school rate curves as smooth along cohort progression and temporally across cohorts. These underlying values then interact with the observed data through data class-specific likelihoods with distinct bias and variance terms that reflect the structural differences between administrative and survey-based data. A particularly notable consequence of the dual setup is that the negative observations are neither excluded nor adjusted, but rather are used to inform the error of all administrative data.

As with any modelling exercise, our approach to out-of-school rate estimation does have limitations. Given the large number of parameters needed to construct a full set of estimates for each country, we find that countries with limited data tend to have substantial uncertainty intervals. As a general rule, we find that our model performs best when there is at least one set of observations every 3 to 5 years, though of course, the more data available, the better the performance. Unfortunately, the countries that tend to miss this minimum also tend to be among those with the most fragile education systems, highlighting the need for more frequent surveys, and additional support for administrative data infrastructure.

As an immediate step to improve education data quality, we do suggest that countries implement internal consistency checks on their data reporting. The problem of negative administrative data is pervasive, affecting countries across all regions and all development stages. While we have taken steps to accommodate such observations and even leverage the information they contain, we do think countries reporting population estimates alongside enrolment counts, even if only in census years, would greatly reduce the frequency of invalid observations, therebey lessening the impacts of the two-source problem, and improving our collect understanding of the variability present in out-of-school rates.

6 Acknowledgements

This paper was authored by Ameer Dharamshi, Bilal Barakat, Manos Antoninis, and Silvia Montoya with considerable input, comments, and feedback from members of the UNESCO Institute for Statistics and Global Education Monitoring Report teams. We would like to thank all of the members of the Technical Cooperation Group on the Indicators for SDG 4 for their input on out-of-school rate estimation. This paper is based on data from the UNESCO Institute for Statistics (UIS), World Population Prospects (WPP), Eurostat, SingStat, the Demographic and Health Surveys (DHS), Multiple Indicator Cluster Surveys (MICS), Integrated Public Use Microdata Series (IPUMS), Encuesta Permanente de Hogares (EPH), Integrated Living Conditions Survey (ILCS), Encuestas de Hogares (EH), Pesquisa Nacional por Amostra de Domicílios (PNAD), Encuesta de Caracterización Socioeconómica Nacional (CASEN), the China Family Panel Studies (CFPS), Encuesta Nacional de Calidad de Vida (ECV), Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU), Encuesta de Hogares de Propositos Multiples (EHPM), Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH), Household and Income Expenditure Survey (HIES), Encuesta Permanente de Hogares Continua

(EPHC), Encuesta Nacional de Hogares (ENAHO), the Russia Longitudinal Monitoring Survey - Higher School of Economics (RLMS-HSE), Household Budget Survey (HBS), and Encuesta Continua de Hogares (ECH) from the years 1999-2019. The responsibility for all conclusions drawn from the data lies entirely with the authors.

7 References

Alexander, Monica, and Leontine Alkema. 2018. "Global Estimation of Neonatal Mortality Using a Bayesian Hierarchical Splines Regression Model." *Demographic Research* 38 (15): 335–72. https://doi.org/10.4054/DemRes.2018.38.15.

Alkema, Leontine, Doris Chou, Daniel Hogan, Sanqian Zhang, Ann-Beth Moller, Alison Gemmill, Doris Ma Fat, et al. 2016. "Global, Regional, and National Levels and Trends in Maternal Mortality Between 1990 and 2015, with Scenario-Based Projections to 2030: A Systematic Analysis by the UN Maternal Mortality Estimation Inter-Agency Group." *The Lancet* 387 (10017): 462–74. https://doi.org/10.1016/S0140-6736(15)00838-7.

Alkema, Leontine, and Jin Rou New. 2014. "Global Estimation of Child Mortality Using a Bayesian b-Spline Bias-Reduction Model." *The Annals of Applied Statistics* 8 (4): 2122–49. https://doi.org/10.1214/14-aoas768.

ArmStat. 2018. "Integrated Living Conditions Survey [Datasets]." National Statistical Service of the Republic of Armenia. https://armstat.am/en/.

Bengtsson, Henrik. 2020a. *Future: Unified Parallel and Distributed Processing in r for Everyone*. https://github.com/HenrikBengtsson/future.

———. 2020b. "A Unifying Framework for Parallel and Distributed Processing in r Using Futures." https://arxiv.org/abs/2008.00553.

Besag, Julian. 1974. "Spatial Interaction and the Statistical Analysis of Lattice Systems." *Journal of the Royal Statistical Society. Series B (Methodological)* 36 (2): 192–236. http://www.jstor.org/stable/2984812.

Carpenter, Bob, Andrew Gelman, Matthew D. Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2017. "Stan: A Probabilistic Programming Language." *Journal of Statistical Software, Articles* 76 (1): 1–32. https://doi.org/10.18637/jss.v076.i01.

Chao, Fengqing, Patrick Gerland, Alex R. Cook, and Leontine Alkema. 2021. "Global Estimation and Scenario-Based Projections of Sex Ratio at Birth and Missing Female Births Using a Bayesian Hierarchical Time Series Mixture Model." *The Annals of Applied Statistics* 15 (3). https://doi.org/10.1214/20-aoas1436.

DANE. 2019. "Encuesta Nacional de Calidad de Vida [Datasets]." Departamento Administrativo Nacional de Estadística. https://www.dane.gov.co/index.php/estadisticas-por-tema/pobreza-y-condiciones-de-vida/calidad-de-vida-ecv.

Dharamshi, Ameer, Bilal F Barakat, Leontine Alkema, and Manos Antoninis. 2022. "A Bayesian Model for Estimating SDG Indicator 4.1.2: School Completion Rates." SocArXiv. https://doi.org/10.31235/osf.io/at368.

DIGESTYC. 2019. "Encuesta de Hogares de Propositos Multiples [Datasets]." Dirección General de Estadística y Censos. http://www.digestyc.gob.sv/index.php/temas/des/ehpm.html.

EFA Global Monitoring Report. 2015. "How Long Will It Take to Achieve Universal Primary and Secondary Education?: Technical Background Note for the Framework for Action on the Post-2015 Education Agenda." UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000233028.

Eurostat. 2000-2021. "Eurostat Demography, Population Stock & Balance (Various) [Datasets]." Eurostat. https://ec.europa.eu/eurostat/web/population-demography/demography-population-stock-balance/database.

Gelman, Andrew, and Donald B. Rubin. 1992. "Inference from Iterative Simulation Using Multiple Sequences." *Statistical Science* 7 (4): 457–72. http://www.jstor.org/stable/2246093.

Hoffman, Matthew D., and Andrew Gelman. 2014. "The No-U-Turn Sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo." *Journal of Machine Learning Research* 15 (47): 1593–623. http://jmlr.org/papers/v15/hoffman14a.html.

HSE. 2000-2019. "Russia Longitudinal Monitoring Survey - Higher School of Economics [Datasets]." National Research University Higher School of Economics; ZAO "Demoscope" together with Carolina Population Center, University of North Carolina at Chapel Hill; the Institute of Sociology RAS. https://rlms-hse.cpc.unc.edu/.

IBGE. 2007-2015. "Pesquisa Nacional Por Amostra de Domicílios [Datasets]." Instituto Brasileiro de Geografia e Estatística. http://observatorio.ministeriodesarrollosocial.gob.cl/index.php.

ICF. 2000-2018. "Demographic and Health Surveys (Various) [Datasets]." Rockville, MD: ICF; Funded by USAID. https://dhsprogram.com/.

INDEC. 2004-2012. "Encuesta Permanente de Hogares [Datasets]." Instituto Nacional de Estadística y Censos de la República Argentina. https://www.indec.gob.ar/indec/web/Institucional-Indec-BasesDeDatos.

Independent Expert Advisory Group on the Data Revolution for Sustainable Development. 2014. "A World That Counts: Mobilising the Data Revolution for Sustainable Development." UN Data Revolution. https://www.undatarevolution.org/wp-content/uploads/2014/12/A-World-That-Counts2.pdf.

INE. 2019a. "Encuesta Continua de Hogares [Datasets]." Instituto Nacional de Estadística. https://www.ine.gub.uy/encuesta-continua-de-hogares1.

———. 2019b. "Encuesta Permanente de Hogares Continua [Datasets]." Instituto Nacional de Estadística. https://www.ine.gov.py/.

———. 2019c. "Encuestas de Hogares [Datasets]." Instituto Nacional de Estadística. https://www.ine.gob.bo/index.php/estadisticas-sociales/vivienda-y-servicios-basicos/encuestas-de-hogares-vivienda/.

INEC. 2018. "Encuesta Nacional de Empleo, Desempleo y Subempleo [Datasets]." Instituto Nacional de Estadística y Censos. https://www.ecuadorencifras.gob.ec/enemdu-trimestral/.

INEGI. 2018. "Encuesta Nacional de Ingresos y Gastos de los Hogares [Datasets]." Instituto Nacional de Estadística, Geografía e Informática. https://www.inegi.org.mx/programas/enigh/nc/2018/.

INEI. 2019. "Encuesta Nacional de Hogares [Datasets]." El Instituto nacional de Estadística e Informática. http://iinei.inei.gob.pe/microdatos/.

Kozyreva, Polina, Mikhail Kosolapov, and Barry M Popkin. 2016. "Data Resource Profile: The Russia Longitudinal Monitoring Survey—Higher School of Economics (RLMS-HSE) Phase II: Monitoring the Economic and Health Situation in Russia, 1994–2013." *International Journal of Epidemiology* 45 (2): 395–401. https://doi.org/10.1093/ije/dyv357.

Landau, William Michael. 2021. "The Targets r Package: A Dynamic Make-Like Function-Oriented Pipeline Toolkit for Reproducibility and High-Performance Computing." *Journal of Open Source Software* 6 (57): 2959. https://doi.org/10.21105/joss.02959.

Leslie, P. H. 1945. "On the Use of Matrices in Certain Population Mathematics." *Biometrika* 33 (3): 183–212. http://www.jstor.org/stable/2332297.

Ministerio de Desarrollo Social y Familia. 2000-2015. "Encuesta de Caracterización Socioeconómica Nacional

[Datasets].” Ministerio de Desarrollo Social y Familia. http://observatorio.ministeriodesarrollosocial.gob.cl/index.php.

Minnesota Population Center. 2020. “Integrated Public Use Microdata Series, International: Version 7.3 [Datasets].” Minneapolis, MN: IPUMS. https://doi.org/10.18128/D020.V7.2.

Neal, Radford M. 2011. “MCMC using Hamiltonian dynamics.” In *Handbook of Markov Chain Monte Carlo*. Chapman; Hall/CRC. https://doi.org/10.1201/b10905.

NSO. 2009. “Household and Income Expenditure Survey [Datasets].” National Statistical Office of Papua New Guinea. https://www.nso.gov.pg/census-surveys/household-and-income-expenditure-survey/.

Peking University Open Research Data. 2010-2014. “China Family Panel Studies [Datasets].” Peking University Open Research Data. https://opendata.pku.edu.cn/dataverse/CFPS?language=en.

SingStat. 2000-2021. “Singapore Residents by Age Group, Ethnic Group and Sex, End June [Datasets].” Singapore Department of Statistics. https://tablebuilder.singstat.gov.sg/table/TS/M810011.

Stan Development Team. 2020a. *Brief Guide to Stan’s Warnings*. https://mc-stan.org/misc/warnings.html.

———. 2020b. *Rstan: R Interface to Stan*. https://CRAN.R-project.org/package=rstan.

The Demographic and Health Surveys Program. 2012. *Demographic and Health Survey Sampling and Household Listing Manual*. ICF International. https://dhsprogram.com/pubs/pdf/DHSM4/DHS6_Sampling_Manual_Sept2012_DHSM4.pdf.

TNBS. 2017. “Household Budget Survey [Datasets].” Tanzania National Bureau of Statistics. https://www.nbs.go.tz/index.php/en/census-surveys/poverty-indicators-statistics/household-budget-survey-hbs.

UIS. 2018. “Out-of-School Children and Youth.” UNESCO Institute for Statistics. 2018. http://uis.unesco.org/en/topic/out-school-children-and-youth.

UNESCO. 2016. “Global Education Monitoring Report 2016 Box 14.2.” UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000245752.

UNESCO Institute for Statistics, Global Education Monitoring Report Team. 2019. “Meeting Commitments: Are Countries on Track to Achieve SDG 4?” UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000369009.

UNESCO Institute of Statistics. 2022a. “Background Information on Education Statistics in the UIS Database.” UNESCO Institute of Statistics. https://tcg.uis.unesco.org/wp-content/uploads/sites/4/2022/03/UIS-Education-Database_Background-Information_EN.pdf.

———. 2022b. “Out-of-School Rate (1 Year Before Primary, Primary Education, Lower Secondary Education, Upper Secondary Education).” UNESCO Institute of Statistics. http://uis.unesco.org/en/glossary-term/out-school-rate-1-year-primary-primary-education-lower-secondary-education-upper.

UNICEF. 1999-2019. “Multiple Indicator Cluster Survey (Various) [Datasets].” UNICEF. https://mics.unicef.org/.

United Nations. 2015. “Transforming Our World: The 2030 Agenda for Sustainable Development.” United Nations. 2015. https://sdgs.un.org/publications/transforming-our-world-2030-agenda-sustainable-development-17981.

UNPD. 2019a. “The 2019 Revision of World Population Prospects.” United Nations. https://population.un.org/wpp/.

———. 2019b. “World Population Prospects 2019: Methodology of the United Nations population estimates and projections.” United Nations. https://population.un.org/wpp/publications/Files/WPP2019_Methodology.pdf.

Vehtari, Aki, Andrew Gelman, Daniel Simpson, Bob Carpenter, and Paul-Christian Bürkner. 2020. “Rank-

Normalization, Folding, and Localization: An Improved $\widehat{R}$ for Assessing Convergence of MCMC." *Bayesian Analysis*, July. https://doi.org/10.1214/20-ba1221.

Wang, Zhengfan, Miranda J. Fix, Lucia Hug, Anu Mishra, Danzhen You, Hannah Blencowe, Jon Wakefield, and Leontine Alkema. 2020. "Estimating the Stillbirth Rate for 195 Countries Using A Bayesian Sparse Regression Model with Temporal Smoothing." arXiv. https://doi.org/10.48550/ARXIV.2010.03551.

Xie, Yu, and Ping Lu. 2015. "The Sampling Design of the China Family Panel Studies (CFPS)." *Chinese Journal of Sociology* 1 (4): 471–84. https://doi.org/10.1177/2057150X15614535.

A Data

A.1 Sources

The out-of-school rate analysis presented in this paper is based on 110 206 out-of-school rate observations from 192 countries.

In total, there are 92 036 administrative observations covering 187 countries. The enrolment component is provided by the UNESCO Institute for Statistics through an internal data sharing arrangement whereas the population data is sourced from the World Population Prospects (WPP) (UNPD 2019a), Eurostat (Eurostat 2000-2021), and SingStat (SingStat 2000-2021).

Survey-based data represents 18 170 observations from a total of 135 countries. The primary survey data sources are the Demographic and Health Surveys (DHS) (ICF 2000-2018), Multiple Indicator Cluster Surveys (MICS) (UNICEF 1999-2019), and census samples extracted from the Integrated Public Use Microdata Series (IPUMS) (Minnesota Population Center 2020). IPUMS extracts were limited to a maximum of one million observations due to computational constraints. In addition, we include the following selection of nationally administered household surveys collected in the World Inequality Database on Education (WIDE):

(a) Argentina: Encuesta Permanente de Hogares (EPH) (INDEC 2004-2012)
(b) Armenia: Integrated Living Conditions Survey (ILCS) (ArmStat 2018)
(c) Bolivia: Encuestas de Hogares (EH) (INE 2019c)
(d) Brazil: Pesquisa Nacional por Amostra de Domicílios (PNAD) (IBGE 2007-2015)
(e) Chile: Encuesta de Caracterización Socioeconómica Nacional (CASEN) (Ministerio de Desarrollo Social y Familia 2000-2015)
(f) China: China Family Panel Studies (CFPS) (Xie and Lu 2015; Peking University Open Research Data 2010-2014)
(g) Colombia: Encuesta Nacional de Calidad de Vida (ECV) (DANE 2019)
(h) Ecuador: Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU) (INEC 2018)
(i) El Salvador: Encuesta de Hogares de Propositos Multiples (EHPM) (DIGESTYC 2019)
(j) Mexico: Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) (INEGI 2018)
(k) Papua New Guinea: Household and Income Expenditure Survey (HIES) (NSO 2009)
(l) Paraguay: Encuesta Permanente de Hogares Continua (EPHC) (INE 2019b)
(m) Peru: Encuesta Nacional de Hogares (ENAHO) (INEI 2019)
(n) Russia: The Russia Longitudinal Monitoring Survey - Higher School of Economics (RLMS-HSE) (Kozyreva, Kosolapov, and Popkin 2016; HSE 2000-2019)

(o) Tanzania: Household Budget Survey (HBS) (TNBS 2017)
(p) Uruguay: Encuesta Continua de Hogares (ECH) (INE 2019a)

A.2 Data Pre-Processing

Preparing observed out-of-school rate estimates for analysis follows a systematic process based on the data source.

A.2.1 Administrative Data

Each year, the UNESCO Institute for Statistics (UIS) receives enrolment data by age, sex, and education level from the Ministries of Education of participating countries. As the definition of out-of-school classifies students in-school whether they are on-time, ahead, or delayed, these enrolment counts must be aggregated across education level for each country-age-sex-year combination. Unfortunately, not all countries provide data in every year and in many cases where enrolment counts are provided, they are incomplete. These incomplete cases must be identified, and in some cases, excluded during the data pre-processing stage. To illustrate why, consider a country with a significant delayed enrolment. Then, if primary enrolment counts are missing, enrolment counts for lower secondary *ages* may be understated even if lower secondary enrolment is present as those that are delayed but are still enroled in primary school are missing. In response to this challenge, we take the following three steps for enrolment count exclusions:

1. First, for a given age, if the enrolment count for the corresponding expected education level is missing (ex. primary counts are missing for primary age children), then the country-age-sex-year in question is excluded.
2. Second, if a given age is the first expected age for a given level but the prior level enrolment count is missing (ex. primary counts are missing for the first age of lower secondary), then the country-age-sex-year in excluded.
3. Finally, in cases where there is a consistently large delayed population, a higher exclusion age threshold in subsequent levels is set. For example, if enrolment counts for age 15 routinely include material levels of primary enrolment, then any age at or below 15 without primary enrolment counts is excluded.

After the exclusion process, total enrolment counts for a given country-age-sex-year, are divided by the corresponding population estimate and subtracted from one to produce out-of-school rate estimates.

A.2.2 Survey Data

Household surveys, including those conducted by MICS, DHS, Census offices, and national statistics offices, are the other source of out-of-school rate estimates. These surveys ask respondents if they attended school during the present school year. These responses are aggregated using the corresponding weights to produce survey-based estimates. Given the sample-based nature of these estimates, sampling errors are computed using the clustered Jackknife procedure as well (The Demographic and Health Surveys Program 2012).

B Model Summary

In this section, we present a complete model summary, including all equations and priors previously discussed in the body of the paper.

First, define A_i as an observed administrative out-of-school rate for country $c[i]$, region $r[i]$, year $y[i]$, and age $a[i]$, and define S_j as an observed survey-based out-of-school rate for country $c[j]$, year $y[j]$, age $a[j]$, survey wave $w[j]$, and survey series $s[j]$. We then model these observations as follows:

$$A_i \sim \mathcal{N}\left(\kappa_{c[i],y[i],a[i]}, \sigma^2 + \omega^2_{r[i]}\right), \tag{16}$$

$$\Phi^{-1}\left(S_j\right) \sim \mathcal{N}\left(\Phi^{-1}\left(\kappa_{c[j],y[j],a[j]}\right) + \beta_{c[j],s[j]}, \frac{\nu_j^2 + \eta^2_{w[j]}}{f\left(\Phi^{-1}\left(S_j\right)\right)^2}\right), \tag{17}$$

where $\kappa_{c,y,a}$ is the underlying true out-of-school rate for country c, year y, and age a, σ is the global administrative data error, ω_r is the regional administrative data error, $\beta_{c,s}$ is the survey bias for country c's surveys originating from survey series s, ν_j is the sampling error for survey observation j, and η_w is the non-sampling variance for survey wave w, and Φ^{-1} refers to the probit function.

The underlying $\kappa_{c,y,a}$ is modeled in the probit space using a cohort structure defined as:

$$\Phi^{-1}\left(\kappa_{c,y,a}\right) = \mu_{c,y-a+1} + \Theta_{c,y,a}, \tag{18}$$

$$\mu_{c,y} = \sum_{k=1}^{K_1}\left(B_k^{(1)}(y)\mu^*_{c,k}\right), \tag{19}$$

$$\mu^*_{c,k+1} = \gamma_c + \mu^*_{c,k} + \epsilon_{c,k}, \tag{20}$$

$$\Theta_{c,y,a} = \sum_{g=1}^{a-1}\sum_{k=1}^{K_5}\left(B_k^{(5)}(y-a+1)\alpha_{c,g,k}\right), \tag{21}$$

where $\mu_{c,y}$ refers to the baseline out-of-school rate for the entry age estimated with a smoothed random walk with drift, γ_c is the country-specific drift term for baseline out-of-school rates, $\Theta_{c,y,a}$ is the changes in out-of-school rate from the baseline out-of-school rate in year $y-a+1$ to the time the cohort reaches age a in year y, and the $\alpha_{c,g,k}$ parameters are the spline coefficients capturing the changes in out-of-school rates.

The γ_c parameter is modeled using a regional hierarchy structure,

$$\gamma_c \sim \mathcal{N}^-\left(0, \sigma^2_{\gamma,r[c]}\right), \tag{22}$$

$$\sigma_{\gamma,r[c]} \sim \mathcal{N}^+\left(0, 0.02^2\right), \tag{23}$$

the $\alpha_{c,g,k}$ values are smoothed across ages and across splines and constrained as,

$$\alpha_{c,g,k} \sim \mathcal{N}\left(\alpha_{c,g-1,k}, \tau_{\alpha_1}^{-1}\right), \tag{24}$$

$$\alpha_{c,g,k} \sim \mathcal{N}\left(\alpha_{c,g,k-1}, \tau_{\alpha_2}^{-1}\right), \tag{25}$$

$$\sum_{g=1}^{a-1} \alpha_{c,g,k} \sim \mathcal{N}\left(0, 2^2\right), \tag{26}$$

and the remaining prior distributions assigned to the process model are,

$$\mu_{c,0}^{*} \sim \mathcal{N}\left(0, 2^2\right), \tag{27}$$

$$\epsilon_{c,k} \sim \mathcal{N}\left(0, \sigma_{\mu}^2\right), \tag{28}$$

$$\sigma_{\mu} \sim \mathcal{N}^{+}\left(0, 0.1^2\right), \tag{29}$$

$$\tau_{\alpha_1} \sim \mathcal{N}^{+}\left(0, 1^2\right), \tag{30}$$

$$\tau_{\alpha_2} \sim \mathcal{N}^{+}\left(0, 1^2\right), \tag{31}$$

where $\mu_{c,0}^{*}$ is the baseline out-of-school rate for the year 2000.

For survey-based data, $\Phi^{-1}\left(\kappa_{c,y,a}\right)$ values are then adjusted for source-specific bias $\beta_{c,s}$ using a hierarchical structure that pools information on bias scaling across countries as follows:

$$\beta_{c,s} \sim \mathcal{N}\left(0, \sigma_{\beta,s}^2\right), \tag{32}$$

$$\sigma_{\beta,s} \sim \mathcal{N}^{+}\left(0, 0.1^2\right). \tag{33}$$

Each observation is subject to an error composed of two components. For administrative data, σ refers to the global error and ω_r is the regional error. For survey data, ν_j is the sampling error estimated prior to modelling using the clustered Jackknife, and η_w is the non-sampling error for each survey wave. The σ, ω_r, and η_w terms are estimated using a shared scaling as follows:

$$\sigma \sim \mathcal{N}^{+}\left(0, \sigma_{\sigma}^2\right), \tag{34}$$

$$\omega_r \sim \mathcal{N}^{+}\left(0, \sigma_{\sigma}^2\right), \tag{35}$$

$$\eta_w \sim \mathcal{N}^{+}\left(0, \sigma_{\sigma}^2\right), \tag{36}$$

$$\sigma_{\sigma} \sim \mathcal{N}^{+}\left(0, 0.01^2\right). \tag{37}$$

For survey-based observations, errors are estimated in the real space to enhance comparability with administrative data and then transformed to the probit space as $\frac{\nu_j^2+\eta_{w[j]}^2}{f(\Phi^{-1}(S_j))^2}$ where Φ^{-1} is the probit function and f is the standard normal pdf.

C Computation

The model is implemented in Stan version 2.21.5 and executed in R version 4.0.2. We run 4 chains per population with each chain consisting of 2 000 burn-in iterations and 2 000 samples before being thinned to a total of 1 000 samples per population. The chains were run in parallel with the assistance of the `future` package (Bengtsson 2020a, 2020b). To enhance reproducibility, the `targets` package was used to manage the modelling pipeline (Landau 2021).