New global estimates of mean years of schooling for 171 countries

Michaela Potančoková, Samir K.C., Anne Goujon, Ramon Bauer

The frequently used indicator of mean years of schooling (MYS) has the advantage of expressing the distribution of educational attainment in a single number. It is therefore often used for cross-country comparisons as well as in economic and environmental models as the unique indicator of educational attainment and human capital stock. The computation of mean years of schooling from a given educational attainment distribution is complex for two main reasons. First, the standard duration of different levels of schooling varies from country to country, and within countries each school level can have different lengths in different studies, for example, studies of general secondary as opposed to vocational secondary. Secondly, the calculation is biased by the presence of pupils/students who do not complete the full course at any level, which can amount to a substantial share in some countries. To contravene these difficulties, the methodology used and detailed in this section computes mean years of schooling as the weighted mean of six educational levels and the procedure takes into account country-specific educational systems as well as changes in these systems over time. We developed regional sets of regression models to improve estimates of MYS for the incomplete primary category and a set of correction factors to adjust higher levels. The models are built using detailed data on duration of schooling by grades completed within primary level for 57 countries (using micro-data from the IPUMS and DHS).

We apply the method to estimate MYS for 171 countries in the WIC dataset on educational attainment as well as to the new set of the Wittgenstein centre human capital projections. The new set of projections draws a global picture of educational attainment levels today and alternative scenarios for their evolution over the rest of the century. Compared to previous work (KC et al., 2010; Lutz et al., 2007), three important changes were implemented regarding data structure and coverage in the current projections: the projection base-year data were updated to the year 2010 instead of 2000, the number of education categories was increased from four to six to encompass a broader range and more variability in levels of attainment, and more countries were added – from 120 to 171 to cover over 97% of world's population in 2010. The harmonised baseline educational attainment compositions by age and sex make the Wittgenstein centre dataset the most comprehensive comparative dataset on educational attainment available (see Bauer et al. 2012 for details).

We also compare our results and method to the widely used Barro and Lee data (Barro and Lee 2013) and explain the differences. These arise due differences in the baseline data as well as methods used to estimate up to date educational attainment and a procedure to compute mean years of schooling.

Method

The methodology used computes mean years of schooling as the weighted mean of six educational levels based on ISCED 1997 classification - no formal education, incomplete primary, completed primary, completed lower secondary, upper secondary and post-secondary education (see Bauer et. al 2012 for detailed definition of education categories) and takes into account country-specific educational systems as well as changes in these systems over time. Information on duration of schooling of completed ISCED levels is taken from the UIS database. For the cohorts that studied prior to the 1970, which is the last year for which UIS provides information, we assume due to lack of

empirical evidence that durations are the same as in the last year of observation. For the projected periods we use durations for 2010.

Although the majority of persons with completed primary, lower secondary or upper secondary level of attainment did not study any further, each of these categories includes a small fraction of individuals who studied some years longer at the next higher level but did not complete it. Since data on the six aggregate levels of education do not include detailed information by levels and grades studied, we needed to estimate duration of schooling for incomplete primary level in each country using correction factors for completed primary, lower secondary and upper secondary education. For post-secondary education, which is a broad category comprising non-university and university education as well as postgraduate education, we apply years of schooling equivalent to four years in all countries with the exception of Singapore, where the post- secondary category consisted of some people with duration equivalent to that in the upper secondary. Age and sex specific proportions of these persons at the post-secondary level were estimated and used to correct the average years of schooling.

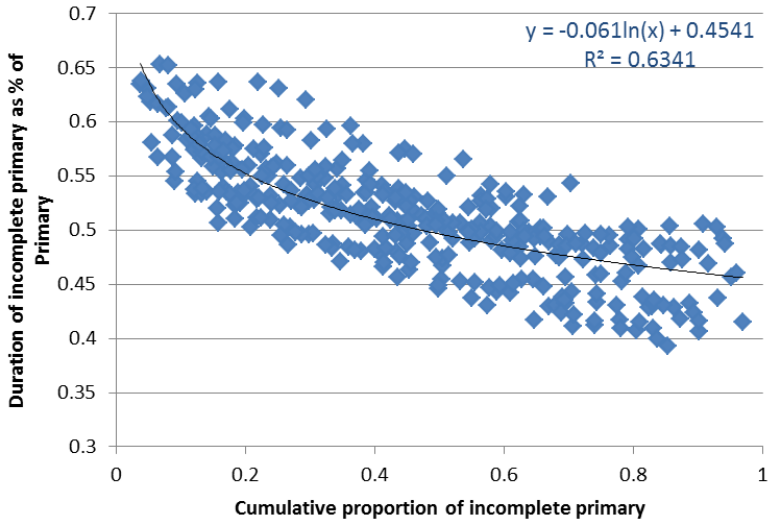
Earlier methods to convert distributions into MYS (Lutz et al. 2007, KC et al. 2010) estimated average durations in these categories using the official definition of the duration weighted by the educational distribution above and below each category. For example, a person with completed primary education might have spent time in school that ranges from the duration of primary to one day less than the duration of lower secondary. An average is obtained from the middle fifty percent of this range, such that the average years of schooling will be larger if proportion below the specific education category is lower compared to the proportion above This method, though intuitive, was found to overestimate average years of schooling when compared with the average years of schooling computed directly from the census micro-data and from surveys.

In our new approach, duration of schooling at the incomplete primary level is estimated using a set of simple models for five broad regions: Latin America, South-East Asia, South Asia, Sub-Saharan Africa and Arab countries. The regional distribution was chosen because of distinct differences in the slopes of the regression function. Trends in education are more similar for countries with a shared history and culture. The models are built using detailed data on duration of schooling by grades completed within primary level for 57 countries (using micro-data from the IPUMS and DHS).We expected to find a relationship between the duration of schooling at primary level and educational attainment composition because we assumed that in countries with low educational attainment and attendance, pupils would be more likely to drop out earlier than in societies with high educational attainment, where drop outs are rather exceptional and would occur at higher grades since children are supported to stay in education longer. The analysis we performed confirmed this hypothesis.

We also found that this assumption holds not only across countries but also across cohorts within individual countries: duration of schooling within the incomplete primary level is shorted for older (less educated) cohorts. Therefore for countries and cohorts with nearly universal primary education, we find higher duration of incomplete primary among the fraction that has dropped out of primary. This relationship holds for both genders. Figure 1 depicts the model for Latin America. The results are similar in the four other broad regions mentioned above. For the regions where the necessary data were not available, like Europe, North America, Australia, Oceania, and the ex-soviet countries in

central Asia,[1] we assume the same relationship as in Latin America, i.e. rather high duration of schooling for those with incomplete primary since these regions benefit from high levels of educational attainment. The fraction of the incomplete primary education category in these regions is negligible overall, even for older cohorts. In the projection, duration of schooling for incomplete primary was calculated using the above relationship for a given (projected) cumulative proportion with incomplete primary.

Figure 1: Relationship between duration of incomplete primary education (ISCED 1) and cumulative proportion of up to incomplete primary by cohorts aged 25-80+ in 16 Latin American countries



For primary, lower and upper secondary levels, we have estimated correction factors to inflate average duration of schooling, to take into account the fraction of persons who enrolled into the next higher level – e.g. in upper secondary education for those who have completed lower secondary education – but did not complete it to the highest grade. The correction factors were estimated for three broad regions – Latin America, Asia and Africa--observing changes across different age groups. Differences between the regions are relatively small and therefore we estimated the correction factors for only three broader regions. For primary level we observe a trend across age groups (Figure 2) which we use to adjust the average duration of primary education by age groups. For example, if standard duration of schooling at primary level for age group 25-29 is six years we apply the correction factor of 1.15 to adjust for the fraction of population with incomplete lower secondary education in this age group. The correction factor declines with increasing age. In the projections, these correction factors were applied to respective cohorts, such that at each step, the youngest cohort has the same correction factor as that of the youngest cohort in the baseline.

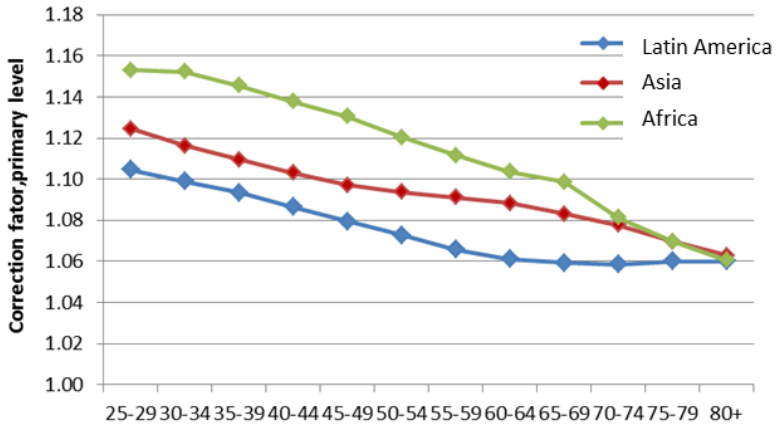
For lower and upper secondary we could not identify any trend and therefore use a single value for all age groups: 1.05 for Latin America, 1.04 for Africa and 1,00 for Asia[2]. For Europe, North America, ex-soviet countries, and Australia and Oceania we use the values for Latin America. For those

[1] Early introduction of universal lower secondary education translated into high completion of this level and negligible proportion of persons with lower educational attainment, which makes these countries distinctly different from other countries in the region.

[2] The value is close to 1 in Asia because most students in countries like India or Nepal, which have educational systems based on the British system, complete 10th grade (ISCED 3C) and only a small fraction completes 12th grade (ISCED 3A).

education categories, the single factor is calculated as the average across age groups, because the variation across ages is small and heaping rather than following a clear trend.

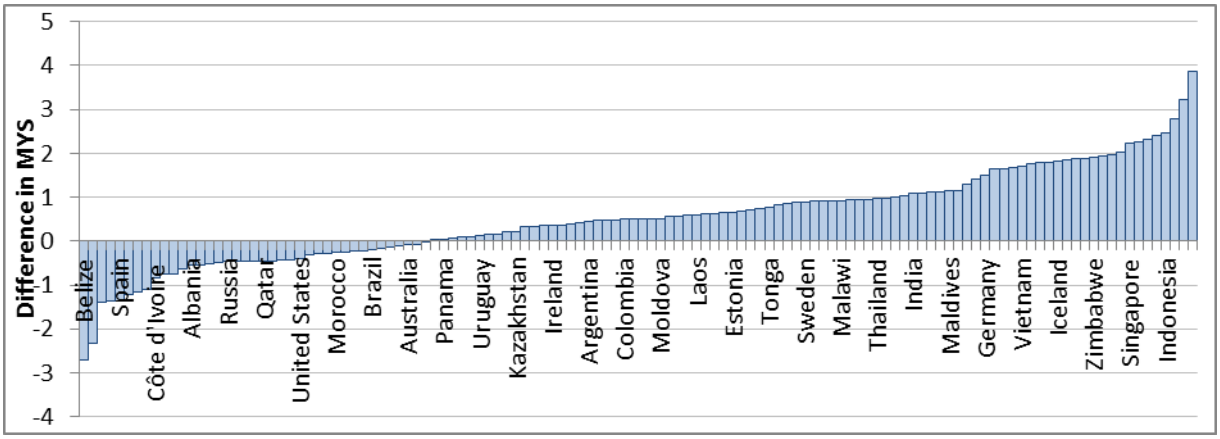
Figure 2: Correction factors for the average duration of completed primary for three broad regions



Results

Results show huge disparities in education across the globe and regions and gaps between genders and across cohorts. In the next step we compare our results to Barro and Lee data which are widely used for international comparisons. We could compare 125 countries, although our dataset contains 46 additional countries. Our procedure leads to comparable results (10% deviation compared to Barro and Lee data) in 54% of the cases with s tendency to provide higher MYS. This is caused mostly by different assumption on MYS equivalent used for post-secondary education and different treatment of the ISCED 4 category. The results section as well as the comparison will be elaborated in more detail in the presentation.

Figure 3: Comparison of the MYS in 2010 from the Wittgenstein dataset and Barro and Lee dataset, 125 comparable countries



References

Barro, R.J., Lee, J.W., 2013. A new data set of educational attainment in the world, 1950–2010. J. Dev. Econ. 104, 184–198.

Bauer, R., Potancokova, M., Goujon, A., K.C., S. 2012. Populations for 171 Countries by Age, Sex, and Level of Education around 2010: Harmonized Estimates of the Baseline Data for the Wittgenstein Centre

Projections

IIASA Interim Report IR-12-016

KC, S., Barakat, B., Goujon, A., Skirbekk, V., Sanderson, W.C., Lutz, W., 2010. Projection of populations by level of educational attainment, age, and sex for 120 countries for 2005-2050. Demogr. Res. 22, 383–472.

Lutz, W., Goujon, A., KC, S., Sanderson, W.C., 2007. Reconstruction of populations by age, sex and level of educational attainment for 120 countries for 1970-2000. Vienna Yearb. Popul. Res. 2007 5, 193–235.