Projecting Global Educational Attainment to 2050 Using a Hierarchical Bayesian Model

submitted to the European Population Conference 2012
Stockholm, 13–16 June 2012

Bilal Fouad Barakat
Vienna Institute of Demography (VID)
bilal.barakat@oeaw.ac.at

Extended Abstract

There has in recent years been an increased interest in applying hierarchical Bayesian models to the problem of projecting trends in demographic parameters, such as TFR, urbanization, or disease prevalence. However, population educational attainment has to date not been investigated in this manner, despite the fact that education is intrinsically a long-term, forward-looking process, and future expansion paths are crucial to assessing the feasibility of international development goals such as Education For All (EFA) or the education-related Millennium Development Goals (MDG).

The approach followed here is similar in spirit to that by (Alkema et al., 2011), but less complex, because the convergence state of universal educational participation is much less volatile than an average TFR level, say. Once primary enrollment is truly universal, say, it does not randomly fluctuate down to 96% in some years. As a result, no switching between different model regimes is necessary. The model below is implemented independently for each level of schooling (the corresponding index is dropped for clarity).

$$\text{participation}_{it} = \Phi(x_{it})$$

$$x_{it} = x_{i(t-1)} + g + r_i + c_i + \epsilon_{it}$$

where Φ is the c.d.f. of a standard Gaussian random variable, and $g, r_i, c_i, \epsilon_{it}$ represent global, regional and country random effects and residuals for country i. In other words, the trajectories follow an s-shaped inverse probit path, and progress at each step is driven by a global trend, regional trends, and country-specific trends, as well as random noise. Because we are not interested in lateral shifts in time, the value of x_0 is not of intrinsic interest and the data can be centered so that participation is 50% at time t_0 for estimation purposes. The resulting participation is to be interpreted as the share of individuals in the relevant age cohort at time t who were previously in the attainment level below, and progress to the higher level. Saturation below 100% is not observed even for tertiary education. Nevertheless, extremely high tertiary participation rates that seem implausible from today's perspective lie beyond the projection horizon.

All of these have diffuse Gaussian priors with their own variances. Estimation is performed using Markov Chain Monte Carlo sampling. The model is fitted to a data-set of reconstructed educational attainment data for 120 countries for the period 1970–2000 (Lutz et al., 2007).

This relatively straightforward model fits the data surprisingly well. In particular, the inverse probit growth path was given preference over a logistic or polynomial spline model during exploratory analysis.

Additional regressors, such as GDP or population growth, were not included, partly because there are few indicators available in form of consistent time series for all the countries in the sample, and more importantly because these effects are implicit in the educational expansion dynamic, and including them explicitly raises serious endogeneity problems. In any case, the ultimate aim of the exercise is projection, so the predicted growth paths are more important than the estimated parameter values themselves.

These projections are — like all projections — wrong. However, depending on how wrong they are, and how much we can know about their error, they might still be useful. There is no way of knowing this without systematically assessing their accuracy, or at least, how far off they are from expectations formed taking into account more information than merely past numeric trends.

The next step is therefore to canvas qualitative and quantitative assessments of the projections from country experts in academia, national ministries or statistical offices, and international organisations. This takes the form of an online survey asking whether the projected trajectory for a given country is deemed too high or too low, if adjusting the regional and global convergence parameters leads to a more plausible projection, and qualitative elaborations on the reasons for the above judgements. An evolving, functional, prototype for this exercise can be found at: `http://www.iiasa.ac.at/Research/POP/EduProjOnlineConsultationTRIAL/`.

This is a major contribution, because there is little to no tradition of long-term quantitative projections in international education, and assumption-setting processes that are routine in demographic applications are little established, and there is a pronounced reluctance to provide numerical estimates. This need to keep things as simple as possible is also the reason for seeking post-hoc evaluations of the resulting projections rather than a priori elicitation of expert views for the formulation of informative prior parameter distributions.

The resulting projections, combining sophisticated statistical extrapolation with expert knowledge, are interesting both on their own terms and because of their implications for research on future human capital and retirement. By way of example, Figures 1 and 2 shows illustrative projection trajectories, for tertiary education of females in Portugal and lower secondary education of females in Turkey respectively. While the spread appears relatively modest, and may suggest the need to include model uncertainty, the variation in growth is actually substantial. The time at which female secondary attainment is projected to cross the threshold of 90% differs by more than 15 years between the 5% and 95% quantiles. Either way, judging by past experience, encoded in the trends, it may take at least two generations to approach universal completion of this level of schooling. An extension to take into account uncertainty of the status quo value is anticipated. The example from Portugal shows that the past expansion of female tertiary participation has been very much faster than the global average, to the extent that the global trend essentially coincides with the lowest 5% quantile projection. The high quantile trajectories clearly shows the underlying s-shaped model.

As a result of the hierarchical model structure, where information is exchanged between all countries, and additionally between countries in a given world region, the projections reflect not only the past experience of the country in question, but of shared trends, even without assuming a convergent trend.

References

Alkema, L. et al. (2011). “Probabilistic projections of the total fertility rate for all countries”. In: *Demography*, pp. 1–25.

Lutz, W. et al. (2007). “Reconstruction of populations by age, sex and level of educational attainment for 120 countries for 1970—2000”. In: *Vienna Yearbook of Population Research*, pp. 193–23.

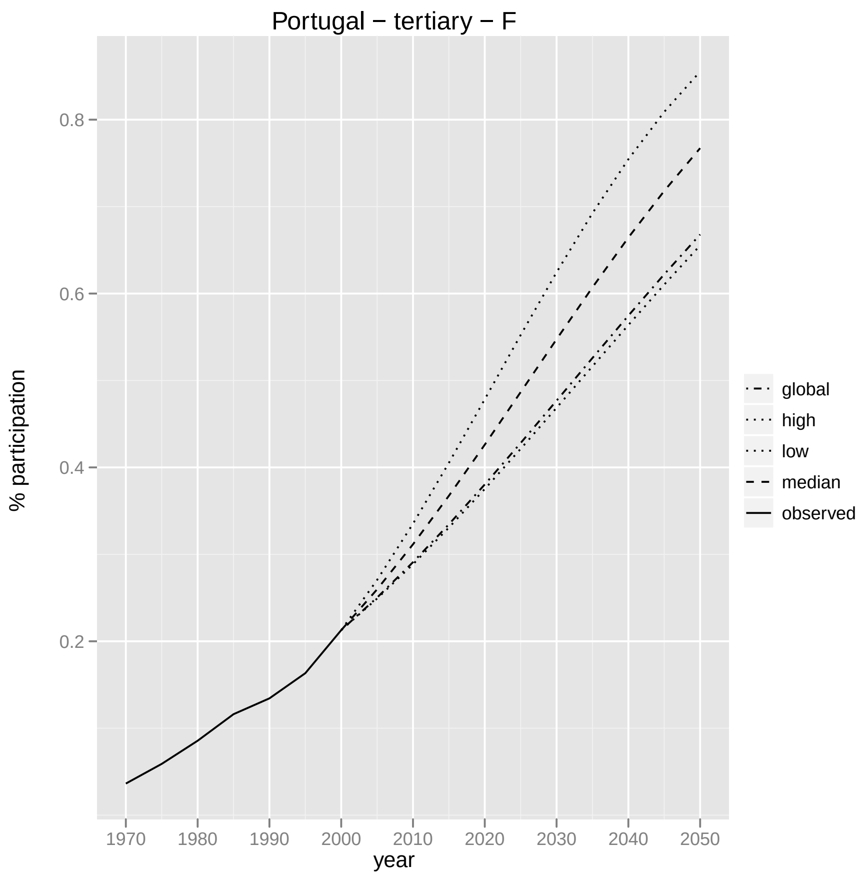

Figure 1: median, 5%, and 95% posterior quantile projections

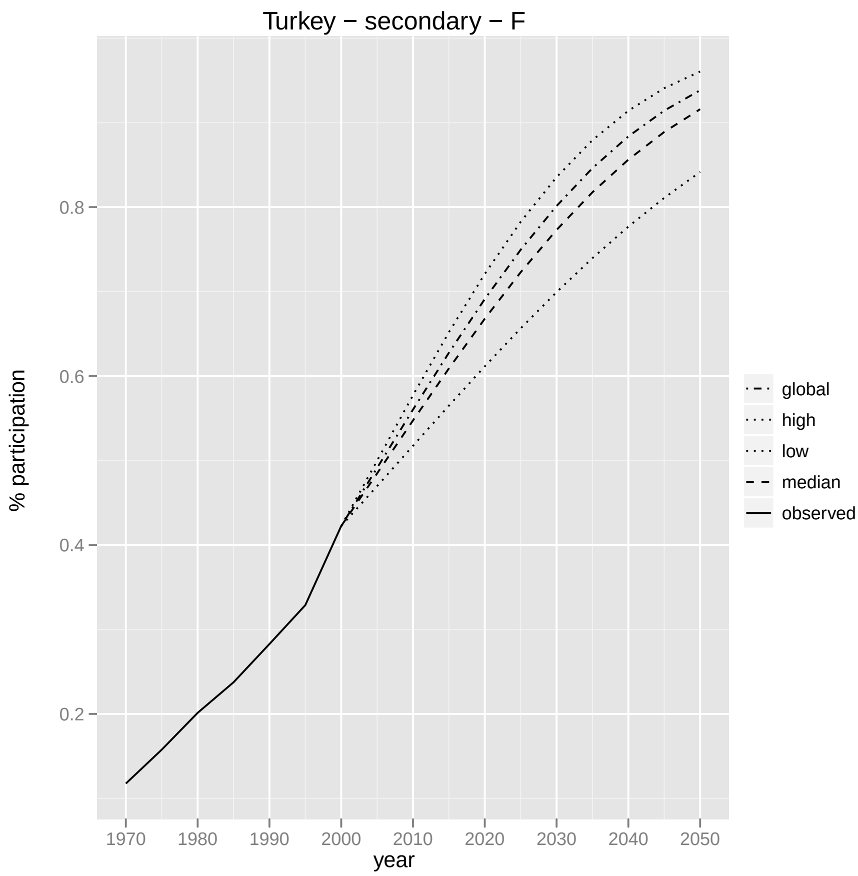

Figure 2: median, 5%, and 95% posterior quantile projections