Analysis and projection of the Human Capital Index using the DYNAMIS-POP micro-simulation model

Martin Spielauer

First draft - Oct 2018

Abstract: This note outlines how the DYNAMIS-POP micro-simulation model can be used for human capital projections and related analysis. The newly introduced World Bank Human Capital Index (HCI) addresses the incentive problem of investments in human capital where returns are slow. HCI is "designed to highlight how investments that improve health and education outcomes today will affect the productivity of future generations". Microsimulation is a specifically powerful tool for the analysis of systems and policies with such a strong longitudinal component. We use DYNAMIS-POP for Nepal highlighting how micro-simulation can inform policy analysis and project the impact of observed socio-demographic and education changes on the HCI.

Introduction

DYNAMIS-POP is a portable dynamic micro-simulation platform allowing to create realistic data-driven simulations with emphasis on applications addressing development issues. It is based on data readily available for most developing countries. As a highly modular system it can be adapted and refined for specific uses in a broad field of policy-relevant applications. DYNAMIS-POP supports the interaction between people and the alignment of aggregated outcomes to existing demographic projections. In this note we present some preliminary analysis for Nepal based on projections starting from the year 2001 based on 2001 Census and DHS data. We have chosen this starting year for allowing retrospective projections supporting model validation.

The HCI is an aggregate measure composed of five components (values for Nepal as released 2018):

- Child survival (cs): the probability of a newborn to survive the first 5 years (0.97).
- Adult survival (as): the probability of a 15-year-old to survive up to age 60 (0.85).
- The stunting rate (sr) of children age 0-4 (0.36).
- The expected years of schooling (ys) including pre-school by the age of 18 (11.7).
- The quality of schooling (qs) for quality adjustment of school duration (0.59)

The HCI is calculated as:

$$HCI = cs * \exp(0.08(ys*qs-14)) * \exp((0.65(as-1)+0.35(-sr))/2)$$

$$HCI = 0.97 * \exp(0.08(11.7*0.59-14)) * \exp((0.65(0.85-1)+0.35(-0.36))/2) = 0.49$$

Besides the quality of schooling (which is added as a parameter) DYNAMIS-POP explicitly models and projects the components of the HCI on a sub-national level typically based on a set of individual level characteristics like sex, mother's education, stunting, and education. While the model can be aligned to aggregate demographic projections, it creates realistic individual life-courses and life-course interactions. For example, mother's education influences child mortality, stunting, and the educational attainment of children. This allows projecting the human capital of a society and how it is driven by composition effects versus trends in individual level outcomes for given characteristics. This is valuable when trying to assess policy effects.

The DYNAMIS-POP model

DYNAMIS-POP produces detailed population projections including educational attainment and school attendance, first marriage, transmission of ethnicity, and a detailed model for infant mortality accounting for mother’s characteristics. The model can be used as is or adapted and extended to meet specific needs. For the following analysis we have added three simple modules complementing the existing modules for the analysis of human capital and the HCI. The new modules add pre-school education, stunting, and the general calculation and table output of the HCI and its components.

- **Survival to age 5:** Child survival is modeled explicitly in DYNAMIS-POP accounting for mother's age at birth and mother's education. (The list of population groups for which relative risks are applied is generic and can be modified and extended easily.) Calibration to national level outcomes is done automatically if this option is selected by the user. Calibration can also be performed just once for the starting year, with future developments being entirely driven by composition changes, i.e. the changing age and education structure of mothers.

- **Adult survival** is modeled by a standard period life table by sex and a parameter for the projected period life expectancy by sex. The latter is used to adjust the standard life table to result in the target life expectancy. The HCI of a cohort is entirely based on the period rates in the year of birth. While DYNAMIS-POP could be used to produce the expected HCI based on the projected period rates resulting from official population projections, we have "frozen" the life expectancy parameters from 2018 onwards thereby reproducing the expected period measures as used in the calculation of the current HCI.

- **Years of schooling:** The HCI accounts for 14 years of school (2 years of pre-school + 12 years of elementary and secondary). DYNAMIS-POP contains a detailed model for primary education (6 grades for Nepal) by mother's education, district, and observed trends. For this analysis we have added stunting as an additional relative factor for deciding school entrance and graduation. Students are tracked through the grade system which delivers measures on the years in school including for school dropouts. Secondary education is implemented by period progression and repetition rates. We have added a pre-school module to add up to 2 school years to the modeled careers. Pre-school years are added at school entrance, assuming that all pre-school children enter school. While the primary school module is quite elaborate producing a realistic selection of students and graduates by individual background, the other modules are simple, and mechanic based on (at this point rather ad-hoc) rates which do not change over time. Consequently, all educational changes produced by the model stem from changes in primary education. The current parameterization - while producing quite accurate projection for the average time in school - is for illustration only and additional data work must be performed for higher accuracy.

- **Adjustment for quality of education:** We apply the aggregate measure as released by the World Bank and only add a term for standard deviation which was chosen ad-hoc.

- **Rate of stunting children blow 5:** We added an individual level module for stunting by district and mother's education based on DHS 2001 data. Stunting is the impaired growth and development that children experience from poor nutrition, repeated infection, and inadequate psychosocial stimulation. Children are defined as stunted if their height-for-age is more than two standard deviations below the WHO Child Growth Standards median. Stunting rates in 2001 were far higher compared to today (about 52%, with a variation by mother's education groups and region from 30% to 63%). We keep the 2001 rates in the simulation as time-fixed parameters, thus all trends in projected stunting rates are due to changes in the education composition of mothers.

DYNAMIS-POP projections

The following illustration departs from the base projection scenario as described above. The projection starts from 2011. The projected HCI for 2018 is 4.75, slightly below the published value 0.49. The difference can be mostly attributed to the improvements in stunting. The models prediction (a drop from above 52% to around 44%) are entirely driven by composition effects by region and mother's education and do not contain any trend; according to data, stunting is around 36%. In consequence, around 50% of the improvement in stunting can be attributed to composition effects.

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Stunting rate	0.517	0.512	0.508	0.501	0.506	0.500	0.493	0.492	0.489	0.483	0.480	0.477	0.469	0.465	0.461	0.455	0.451	0.444	0.443	0.438
Child survival rate	0.924	0.925	0.929	0.933	0.934	0.940	0.943	0.944	0.945	0.950	0.949	0.954	0.955	0.955	0.958	0.959	0.961	0.962	0.961	0.963
Adult survival rate 15-60	0.841	0.843	0.842	0.847	0.845	0.839	0.841	0.848	0.843	0.845	0.845	0.849	0.844	0.841	0.840	0.841	0.843	0.845	0.847	0.847
Average quality of schooling	0.601	0.600	0.600	0.600	0.601	0.598	0.601	0.601	0.600	0.601	0.600	0.601	0.599	0.599	0.599	0.600	0.599	0.599	0.599	0.600
Average years of schooling	10.347	10.390	10.519	10.641	10.715	10.804	10.869	10.968	10.929	11.036	11.074	11.158	11.202	11.222	11.293	11.361	11.379	11.382	11.437	11.475
Average individual level HCI	0.446	0.447	0.452	0.457	0.458	0.461	0.466	0.469	0.468	0.474	0.474	0.479	0.479	0.480	0.483	0.485	0.487	0.489	0.490	0.492
HCI	0.430	0.432	0.436	0.442	0.444	0.447	0.452	0.456	0.455	0.460	0.461	0.466	0.467	0.467	0.470	0.473	0.475	0.476	0.478	0.480

Table: DYNAMIS-POP projections Nepal, Base 2001 -- Base Scenario

Besides this down-stream effect, education enters the HCI also directly. While in the simulation scenario we kept quality of schooling constant, projected average years of schooling increase from 10.3 to 11.4. Again, DYNAMIS-POP can be used to decompose this increase into a composition effect (by mothers' education) and other factors. For doing so, we run an alternative scenario which keeps the probabilities to enter and graduate from primary education constant for given sex, district and mother's education. Such a scenario can be interpreted as a status quo scenario from a mother's perspective.

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Stunting rate	0.5171	0.5118	0.5078	0.5007	0.5059	0.5	0.4928	0.4919	0.4895	0.4825	0.4797	0.477	0.4668	0.4623	0.461	0.4559	0.4536	0.4464	0.4439	0.4344
Child survival rate	0.9239	0.9248	0.9291	0.9328	0.9343	0.9398	0.9428	0.9436	0.9454	0.95	0.9485	0.9532	0.9534	0.9537	0.9584	0.9606	0.9616	0.9618	0.9634	0.963
Adult survival rate 15-60	0.8404	0.8474	0.8462	0.8422	0.8445	0.8438	0.8429	0.8443	0.8462	0.8439	0.8439	0.8462	0.8462	0.8398	0.8457	0.8435	0.8408	0.8466	0.843	0.842
Average quality of schooling	0.6011	0.6002	0.6009	0.5994	0.5994	0.6005	0.5998	0.5992	0.6001	0.6004	0.6005	0.5996	0.5995	0.5994	0.599	0.5999	0.6002	0.6004	0.5993	0.5982
Average years of schooling	10.257	10.419	10.443	10.442	10.512	10.527	10.551	10.593	10.609	10.686	10.671	10.731	10.725	10.758	10.794	10.827	10.855	10.849	10.928	10.935
Average individual level HCI	0.4426	0.4481	0.4513	0.4522	0.4543	0.4581	0.4601	0.4614	0.4633	0.4672	0.4667	0.4706	0.4712	0.4711	0.4747	0.4771	0.4784	0.4795	0.4809	0.4808
HCI	0.428	0.433	0.436	0.437	0.439	0.443	0.445	0.446	0.448	0.452	0.452	0.455	0.456	0.456	0.460	0.462	0.463	0.465	0.467	0.467

Table: DYNAMIS-POP projections Nepal, Base 2001 -- Alternative Scenario without education trends on the micro level.

In this scenario, composition effects increase the average years of schooling from 10.3 to about 10.85. Compared to the projected improvement to 11.4 years, again 50% of the improvement can be attributed to composition effects.

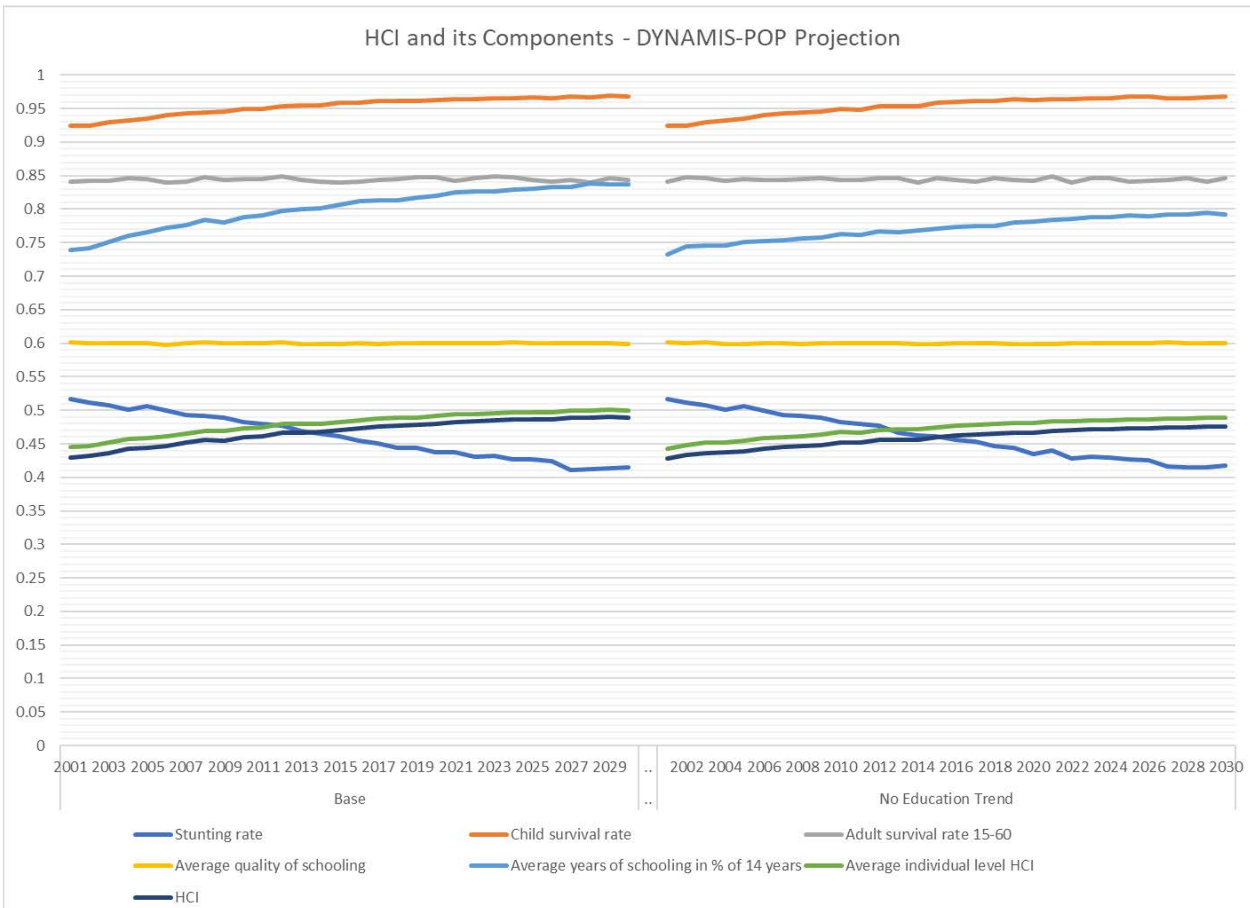

Figure 1: DYNAMIS-POP projections Nepal, Base 2001 -- Base Scenario versus a Scenario without education trends on the micro level.

The same decomposition can be performed for child survival. DYNAMIS-POP includes relative risks by age of the mother and be mother's education and we can switch off other trends.

Discussion

The described simulation exercise demonstrates some contributions DYNAMIS-POP can make in the analysis related to the Human Capital Index.

- DYNAMIS-POP allows to create what-if projections of the HCI and its components. Of particular interest (as benchmark for the assessment of policy effects) are status quo scenarios on the micro level. In such scenarios, all trends on the macro level result from a changing population composition, while for people belonging to a specific group (e.g by region, education, ethnicity) the factors entering HCI (mortality, stunting, education) stay unchanged.
- Other 'popular' what-if scenarios address the differences between population groups and policies targeting specific vulnerable groups. For example, we might assess the impact of narrowing the education gap between population groups on the HCI within this group and on the population level in the long run.
- DYNAMIS-POP allows a decomposition analysis of historic and projected changes in the HCI. This analysis goes beyond the contribution of the individual components of the HCI, as we can assess which part of the changes within each component stem from composition effects.
- DYNAMIS-POP operates on a regional level allowing to assess regional variations in the HCI and its components.
- DYNAMIS-POP allows to quantify the downstream effects of policies. For example, improvements in education influence the HCI not only directly (e.g. more years in school) but also impact child mortality and stunting. Microsimulation can measure these effects and project the timeline of improvements.

Predictions on the population level

The HCI is a cohort measure and does not allow to directly infer on the future average human capital of the working age population. For example, high child mortality substantially lowers the HCI of a cohort but does not impact the human capital of the survives who make the future work force.

Also, as a macro index, the HCI is constructed by a formula linking together the population averages of the components of the measure. As the components are highly correlated on the micro level (e.g. stunting increasing child mortality, lowering education prospects etc.) the HCI is not identical with the average human capital of the population. DYNAMIS-POP allows comparing the HCI with the average human capital of the simulated cohort, the latter being higher, as depicted in Figure 1. (The individual human capital can be calculated in the simulation applying the same formula at the death of each simulated person). Being able to make statements on the average individual level human capital and its distribution might be valuable in economic modeling. In general, one of the strengths of DYNAMIS-POP is its ability to simultaneously produce various alternative measures and their distribution, which can then be picked according to the concrete research question.

As also shown in the example, micro-simulation can be used to retrospectively project the human capital of cohorts born in the past. Such historic simulation might be useful to impute human capital measures into the current population and assess the human capital and its evolvement over time of the total working-age population rather than single cohorts. Again, microsimulation can not only show the effect of a change (or policies; like education improvements) on the population level, but also produces the timeline how changes today impact the future society.

Notes on potential improvements of DYNAMIS-POP for HCI Analysis – for discussion

Mortality:

- Separation of child mortality and adult mortality. Currently we have life expectancy as parameter which scales the mortality table. Child mortality comes on top and includes added risks (mothers age, education) for children 0-4. Child mortality can be calibrated automatically to the overall rates for an initial year and it can be chosen if macro trends (calculated from changing life expectancy) or specific trends for child mortality are added. It would be more logical to have adult life expectancy parameterized separately, e.g. adult life expectancy instead of life expectance at birth.

- In connection, also adult mortality could include relative risks. There is quite some literature e.g. life expectancy differentials by education; (something on stunting?) This would be valuable for benchmark scenarios finding out how much improvement in mortality can be expected from the changing education & stunting composition alone.

Education:

- I like our approach focusing on primary education when it comes to relative risks etc. It is quite robust working from census data capering current trends, working on the district level, and being able to run scenarios driven by composition effects only. Still needs data & literature work for getting relative risks by stunting.

- The existing secondary model might not be so unreasonable for our type of analysis (but requires data work too!). It keeps the model simple but understandable. It assumes that all parental and health influences only apply to primary; when keeping intake and progression rates constant, all improvements are driven by primary school, and more primary graduates result in proportionally more secondary students and graduates.

- Test scores. This would be an interesting area for modeling, also methodological (multi-level analysis was developed in this field). There is if I remember right a high correlation of pre-school experience with test scores, which makes modeling pre-school more interesting and relevant too. This might be a nice project with country partners in this field of research.

Stunting:

- Have to start with more data work. There was a change in definition between DHS 2001 and 2011. Somehow, we will have to model time trends based on some research... maybe this could make a nice collaborative project with experts too...

General:

- I would suggest discussing and developing a set of standard benchmark simulations which project the trends in the HCI
    - E.g. in a “micro steady state world”, where changes in the HCI are driven by composition effects by education, region and ethnicity only. (Only improvements beyond that improve things from an individual’s perspective; might be a benchmark for assessing policy effects which are not pure composition effects)
    - E.g. (the same scenario) but including demographic trends as in official population projections.
    - E.g. (the same scenario) but reaching universal primary education in a decade.
- Any chance to do a country pilot / collaborative projects in developing and finetuning modules and data work?