



OPEN

DATA DESCRIPTOR

A dataset of human capital-weighted population estimates for 185 countries from 1970 to 2100

Guillaume Marois^{1,2} [✉], Jesus Crespo Cuaresma^{2,3}, Jakob Zellmann³ & Claudia Reiter^{2,4}

We provide a novel dataset of human capital-weighted population size (HCWP) for 185 countries from 1970 to 2100. HCWP summarizes a population's productive capacity and human capital heterogeneity in a single metric, enabling comparisons across countries and over time. The weights are derived from Mincerian earnings functions applied to multi-country census data on educational attainment. The model used to compute the returns to schooling accounts for the diminishing positive relative relationship between education and wages as the overall education of populations rises. The population weights are adjusted by a skills assessment factor representing differences in education quality across countries and years. HCWP is calculated by applying these adjusted human capital weights to population estimates and projections disaggregated by age, sex and education, spanning the period 1970–2020 and 2020–2100 for five Shared Socioeconomic Pathway scenarios. Validation analyses demonstrate the utility of the new HCWP data in explaining national income trends. As a more comprehensive population measure than basic size and age-sex indicators, HCWP enhances the power of statistical models aimed at the assessment of socioeconomic change impacts and forecasting.

Background & Summary

This dataset provides human capital weighted population (HCWP) estimates for 185 countries at 5-year intervals, spanning 1970–2020 historically and projected for 2020–2100 under five Shared Socioeconomic Pathways scenarios (SSP1–SSP5)¹. The HCWP adjusts the working-age population (20–64 years) by weights based on educational attainment and education system quality. It summarizes heterogeneity in human capital in a single indicator. Potential uses include assessing past and future economic impacts of demographic and societal changes, comparing productive capacity across countries and historical periods, and predicting income growth making use of econometric models.

Traditional demographic analyses focus on population size and age-sex structure. Age structure, in particular, is an important economic growth predictor. In the decades after a fertility fall, a demographic dividend can emerge, characterized by a high working-age to dependent population ratio. This period tends to see rapid industrialization, education expansion, and economic growth^{2–4}. After some decades under this regime, the elderly share of the population begins rising, bringing population aging and its challenges. The standard narrative is that a higher elderly percentage has negative economic consequences: the retired elderly population produce less but require pensions and healthcare, while the working-age share that provides support and tax revenue shrinks.

Focusing exclusively on changes in the age structure as a predictor of economic growth trends does not provide a complete picture. The ability of the 'supporting population' (the working-age) to actually support the 'dependent population' depends to a large extent on its education, labour force status and productivity, all of which are also undergoing drastic changes. Econometric models estimated on historical data have shown that the inclusion of information on education and age structure, as well as their interplay, contribute significantly to the explanatory power of statistical specifications aimed at reproducing differences in past income and income

¹Asian Demographic Research Institute, School of Sociology and Political Sciences, Shanghai University, Shanghai, China. ²International Institute for Applied Systems Analysis, Wittgenstein Centre for Demography and Global Human Capital (IIASA, OeAW, Wittgenstein Centre for Demography and Global Human Capital (IIASA, OeAW, University of Vienna), Laxenburg, Austria. ³Department of Economics, Vienna University of Economics and Business (WU), Vienna, Austria. ⁴Education and Employment research group, Institute for Advanced Studies (IHS), Vienna, Austria. ✉e-mail: marois@iiasa.ac.at

per capita trends across countries of the world. In particular, the empirical results indicate that including variables which account for age-structured educational attainment levels in economic growth models are better able to account for innovation and technology adoption phenomena^{5–8}.

The Wittgenstein Centre (WIC) Population Projections provide standardized estimates and projections of population by educational attainment for all countries, spanning 1970–2100. Since working-age individuals have varying productivity, these projections use education as a proxy for human capital to enable more refined analysis of demographic change impacts^{8–10}. Education correlates not just with fertility and mortality, but also labor force participation and productivity^{11–16}. For example, WIC projections show India will have a larger working-age population than China in coming decades, but with much lower educational attainment and therefore less productive workers¹⁷. While only 13% of the working age population in China had post-secondary attainment in 2015, versus 34% in the US, this gap is projected to narrow over time as older, less-educated Chinese cohorts are replaced by more educated ones.

The role played by education as a catalyst of socioeconomic change is not only a matter of quantity but also modulated by education quality. There is ample evidence of differences in the quality of education systems among countries in the world¹⁸. A high school diploma does not, on average, provide the same human capital, and its effect on productivity may depend on whether it was obtained in a rich country that spends a lot on education or in a low-income country where schools are underfunded¹⁹.

The human capital-weighted population (HCWP) estimates presented here are calculated by applying weights to the working-age population estimates by educational attainment, which are adjusted by a factor accounting for education system quality. Following an approach similar to other studies^{20,21}, education-specific weights are calculated utilizing a Mincerian earnings function²² on pooled data from all IPUMS-I censuses containing education and income information. The education parameters are interacted with the countries' average educational attainment to account for the dependence of returns to education on the number of workers sharing that education level. Country and time specific adjustment factors for education quality are derived from skills assessment surveys. These final adjusted weights are then applied to population estimates and projections from the Wittgenstein Centre Data Explorer, spanning 1970–2015 historically and projected to 2100 under five Shared Socioeconomic Pathways scenarios (SSP1–SSP5).

We validate our estimates making use of an out-of-sample prediction exercise for GDP per capita growth applied to a panel dataset covering all countries of the world for which data are available. In this exercise, we estimate alternative regression specifications for the growth rate of GDP per capita based on Crespo Cuaresma⁵ and compute several prediction error statistics for the period 2000–2015 (in 5-year steps). In particular, we entertain three different econometric models: a benchmark specification where labor input is measured based on population figures, a second one which includes population by educational attainment level and a third model based on HCWP. Our results indicate that the predictive ability of the model based on HCWP is similar to that of the model based on educational attainment when measured using the global sample, and that these two specifications present a better forecasting performance than the benchmark model.

Methods

Estimates of education-specific weights. *Data sources and selection of the sample.* In order to derive generalizable weights linked to educational levels, we compiled all available censuses from the IPUMS-International database that included both total personal income as well as education variables. We opted to utilize total income rather than wage income alone due to the lack of wage data across many of the surveys. Given that for the majority of individuals, wages typically constitute the primary component of overall income, and the two measures are therefore highly correlated, we make the assumption that total income can serve as a reasonable proxy for wages in our analysis.

Since Mincerian wage regression models assume a log-linear relationship between wages and years of schooling, those persons not active in the formal labor force as well as those with reported incomes at or below zero were assigned an income value equal to one for modeling purposes. Employed individuals with missing income data (4.3%) were omitted from the sample altogether. The IPUMS-International census data leveraged offers the benefit of standardized definitions and categorical binning of key variables. However, the income figures are expressed in national currency units corresponding to the specific period of each survey. As such, we additionally normalized respondents' incomes by dividing through by the national average income level in order to render the values more readily comparable across countries and time.

The education variable is provided in two formats - number of years of education or highest educational attainment level, with varying category breadth. We opted to use years of education in our models since broad attainment categories can obscure substantial within-group heterogeneity. For example, in countries with mandatory education, the lowest group comprises those lacking a high school diploma, while in developing countries surveys more typically differentiate individuals with no schooling from those with some primary or lower secondary schooling.

For surveys only reporting educational attainment, levels were converted to estimated years of schooling as follows:

- Less than primary completed → 1 year;
- Primary completed → 6 years;
- Lower secondary complete → 9 years;
- Secondary completed → 12 years;
- University completed → 16 years.

Country	Year	Total
Canada	1971	14,107
Canada	1981	36,089
Canada	1991	60,552
Canada	2001	67,358
Canada	2011	78,992
Colombia	1973	63,321
Dominican Republic	1981	18,836
Dominican Republic	2002	39,450
Mauritius	2000	42,657
Mexico	1995	48,188
Panama	1980	21,516
Panama	1990	23,340
Panama	2010	23,842
Puerto Rico	1990	47,700
Puerto Rico	2000	52,336
Puerto Rico	2005	25,586
Puerto Rico	2010	26,416
South Africa	1996	20,606
South Africa	2001	23,427
South Africa	2007	27,417
South Africa	2011	30,203
Trinidad and Tobago	1970	28,488
Trinidad and Tobago	2000	69,792
United States	1960	118,991
United States	1970	136,357
United States	1980	64,923
United States	1990	88,255
United States	2000	99,859
United States	2005	102,436
United States	2010	111,243
United States	2015	116,360
Total		1,728,643

Table 1. Selected censuses and their sample size.

Surveys were excluded if attainment categories were insufficiently detailed for comparison to other data sources, such as a single category encompassing both high school and university completion.

We excluded individuals aged under 20 and over 64 from the sample. With the pooled censuses still comprising several million cases even after applying age criteria, we randomly selected a large enough sample from each to generate stable and robust estimates while enabling feasible computational runtimes for the regression models. The included surveys and their corresponding sample sizes are listed in Table 1.

Models. We estimated the education-specific weights from Mincerian earnings function²², which is widely used in the literature to measure the return on investment in education²³. We start by building a simple log-linear model with gamma distribution, predicting the natural log of normalized income (NOR_INC) from years of education (NBEDU), controlling for experience and sex, but ignoring sample design,

$$\ln(\text{NOR_INC}) = \beta_0 + \beta_1 \text{NBEDU} + \beta_2 \text{EXPERIENCE} + \beta_3 \text{EXPERIENCE}^2 + \beta_4 \text{SEX} + \varepsilon \quad (1)$$

β_1 thus provides the semi-elasticity of income to years of education. The variable EXPERIENCE is derived by subtracting from the age the number of years of education and the age of entry at school (set by default to 6 for everyone). Since we aim at estimating generalizable weights, a sample of pooled censuses from different years and different countries is used. Therefore, to account for this sample design, we built a multilevel random-effect model (model 2) which allows the intercept β_0 to vary across censuses (j).

$$\ln(\text{NOR_INC}) = \beta_{0j} + \beta_1 \text{NBEDU} + \beta_2 \text{EXPERIENCE} + \beta_3 \text{EXPERIENCE}^2 + \beta_4 \text{SEX} + \varepsilon_j \quad (2)$$

Finally, we build a final third model to account for the possible variation in the impact of years of education across countries based on the average level of education of their population²³. This model adds parameters β_5 and β_6 for the interaction between the number of years of education at the individual level and a country-level

	Model (1)	Model (2)	Model (3)
Random effect (group = census)		<i>Covariance</i>	<i>Covariance</i>
Intercept		0.208 (0.054)	0.202 (0.054)
	<i>Estimate</i>	<i>Estimate</i>	<i>Estimate</i>
Intercept	-3.257 (0.008)	-3.410 (0.082)	-3.350 (0.337)
NBEDU	0.135 (0.001)	0.146 (0.000)	0.213 (0.002)
NBEDU_MEAN			-0.011 (0.033)
NBEDU*NBEDU_MEAN			-0.006 (0.000)
EXPERIENCE	0.076 (0.000)	0.077 (0.000)	0.077 (0.000)
EXPERIENCE ²	-0.001 (0.000)	-0.001 (0.000)	-0.001 (0.000)
SEX = Male	0.798 (0.003)	0.794 (0.002)	0.794 (0.002)
N	1377624	1377624	1377624
Clusters		31	31
ICC		<0.001	<0.001

Table 2. Parameters of models (standard error in parentheses).

variable that refers to the average years of schooling for the population aged 25 to 54 (NBEDU_MEAN) in each census year.

$$\ln(NOR_INC) = \beta_{0j} + \beta_1 NBEDU + \beta_2 * EXPERIENCE + \beta_3 EXPERIENCE^2 + \beta_4 SEX + \beta_5 NBEDU_MEAN + \beta_6 NBEDU \times NBEDU_MEAN + \varepsilon_j \quad (3)$$

Parameters. Table 2 shows the parameter estimates for the three models employed. Despite income normalization, the covariance estimates in the model given by Eq. (2) indicate statistically significant intercept variability across censuses, evidence supporting the mixed model. The introduction of an interaction term between an individual's years of schooling and the country's average years of schooling in the model given by Eq. (3) reveals that the positive relationship between education and wages diminishes in size as a population's overall education level rises. Specifically, the negative parameter predicts smaller wage gains for marginal increases in an individual's schooling within societies where higher education is more widespread. This is in line with economic principles, as the marginal value of educational attainment decreases when advanced skills become common, and thus the advantages conferred by an extra year of schooling are attenuated when competing against a highly educated population. Thus, the model given by Eq. (3) provides evidence that the labor market return of education depends on both relative and absolute skills and declines in contexts where human capital is abundant. Across all model specifications, the estimated parameters for NBEDU (years of schooling) remain remarkably consistent and comparable, underscoring the reliability of these coefficients.

Using parameters from the model in Eq. (3), we calculate relative weights (W) for each educational level (e), which vary depending on the population's average years of schooling (NBEDU_MEAN) of the country (c) in a given year (t). We postulate that having less than primary completed (e1) corresponds to 1 year of schooling (NBEDU = 1), primary completed (e2) to 6 years, lower secondary (e3) to 9 years, upper secondary (e4) to 12 years and postsecondary (e5) to 16 years. We normalize the weights making use of the value for the world average number of years of schooling in 2015 (9.3). The calculation of W is thus given by

$$W_{e,c,t} = \frac{\exp(\beta_1 * NBEDU_{e,c,t} + \beta_5 * NBEDU_MEAN_{c,t} + \beta_6 * NBEDU * NBEDU_MEAN_{c,t})}{\exp(\beta_1 * 9.3 + \beta_5 * NBEDU_MEAN_{c,t} + \beta_6 * 9.3 * NBEDU_MEAN_{c,t})} \quad (4)$$

Figure 1 below presents weights by educational attainment and average years of schooling. In a country such as the US in 2020, where the average number of years of education for the population aged 25–54 was 13.2 years, a person with postsecondary education would thus be weighted 2.45 times more than a worker with the world average number of years of education and, assuming that income is a good proxy for productivity²⁴, would thus be 2.45 times more productive.

Factorizing for the quality of education. Lutz *et al.*¹⁸ introduced the Skills in Literacy Adjusted Mean Years of Schooling (SLAMYS) to assess the quality of education and human capital. This indicator multiplies the mean years of schooling (MYS) by a factor that takes into account the quality of education (skill adjusted factor or SAF). SAFs are empirically derived from the scores of adult literacy assessments (for individuals aged 20–64) from surveys such as the International Adult Literacy Survey (IALS), the Program for the International Assessment of Adult Competencies (PIAAC), the Skills toward Employment and Productivity Survey (STEP), and the Demographic and Health Survey (DHS). They are calculated cross-sectionally for 185 countries for the period 1970–2015 and normalised to the score for the population-weighted average of OECD countries in 2015 (taken as unity).

To adjust our education-weights for the quality of the education, we use these country- and period-specific SAF from 1970 to 2015 and perform a logit extrapolation for 2015–2100, with the maximum value being the

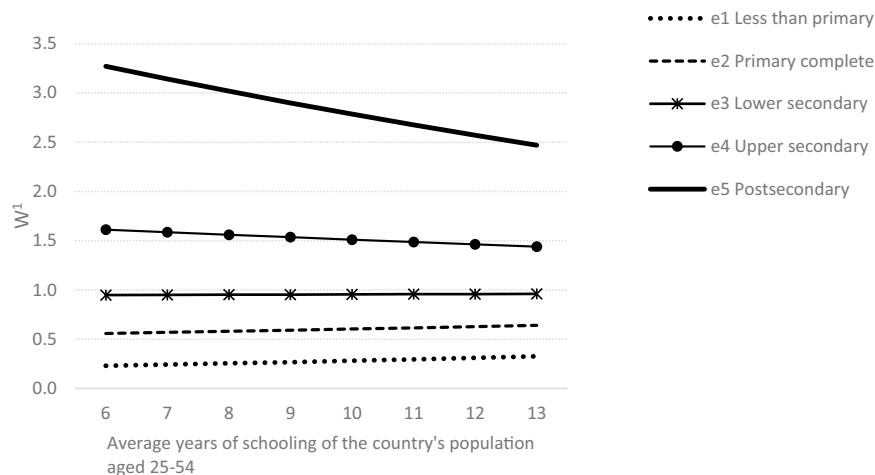


Fig. 1 Weights (W) by educational levels and average years of schooling of the 25–54. $1 =$ world average level of education = 9.3.

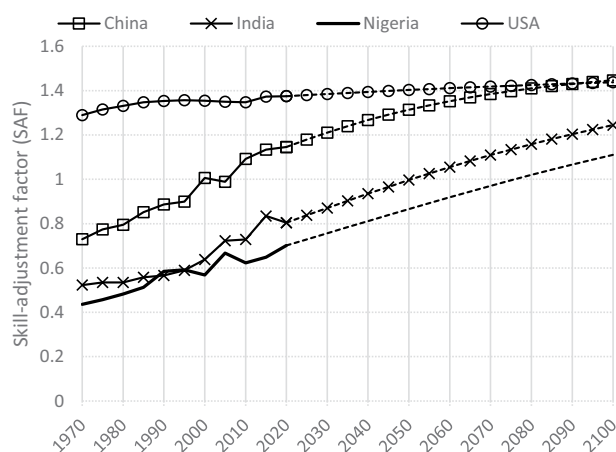


Fig. 2 Skill-adjustment factor ($1 =$ world average in 2015) for China, India, Nigeria and the US, estimated (1970–2015) and projected (2020–2100).

world's highest estimate in 2015 (which is 1.13 for Japan), thus leading to a gradual convergence towards the leading country. We then normalize the score to the world average of 2015. Figure 2 shows the estimated SAF for China, India, Nigeria and the US and the projections from 2020 to 2100.

SAF estimates are cross-sectional, without disaggregation by age, cohort, or specific level of education. In other words, using these values implies to assume no difference in the quality of education across the different degrees within a country. It is also assumed that the change in the cohort size of the population aged 20–64 has only a marginal effect on the trends. Finally, we use the simplifying assumption that the quality of education does not differ between those in the labor force and those out of the labor force.

Population estimated and projected by age, sex, and education. The calculation of projected HCWP requires inputs from population projections by age, sex, and education. We used the most recent update of the population estimates (middle-of-the-road) for the period 1970–2015 and their population projection under five Shared Socioeconomic Pathways scenarios²⁵. The last update of the projections and estimates take into account the impact of COVID-19 on the population structure as well as the latest estimates of demographic components as starting point. The Shared Socioeconomic Pathways scenarios for 2020–2100 are built around narratives and storyline describing alternative socio-economic developments. They were developed by climate scientists and economists to be used as a basis for integrated assessments and future climate modeling. The five scenarios are:

- SSP-1: Sustainability;
- SSP-2: Middle-of-the-road;
- SSP-3: Regional rivalry;
- SSP-4: Inequality;
- SSP-5: Fossil-Fueled Development.

The SSP-2 (Middle-of-the road) scenario can be considered as the reference scenario, as its assumptions are built based on historical patterns. For the mortality assumptions, 75 experts from 30 countries were asked for their opinion on a series of statements on past and future determinants of health and mortality. They concluded that there is room for improvement and that upward trends will continue, with convergence between countries. Country-specific trends are therefore extrapolated with a regional convergence process. Differentials by educational attainment are based on a generalization of estimates where such data are available. For the assumption on international migration, the scenario assumes a continuation of the in and out-migration rates of 1950–2010.

For education, final educational attainment (attained at age 30–34) is extrapolated by gender from country-specific estimates for 1970–2010 with a long-term world convergence. For those with a high level of education, lower levels are imputed at younger ages based on the country's high school graduation age. In this projection, the educational attainment variable is categorical rather than in years of schooling. Therefore, education-weights are calculated based on the normal number of years of education for each level (see data source and sample selection). The future age- and education-specific fertility rates are determined on the basis of a large expert survey in the field of fertility studies^{26,27}. The experts assumed a trend towards global convergence in the very long term. More details on the long-term assumptions and in the scenario definition can be found in Lutz *et al.*^{17,28}.

Calculation of the human capital-weighted population. For a country c at time t , the human capital-weighted population (HCWP) is calculated as:

$$HCWP_{c,t} = SAF_{c,t} \sum_{e=1}^k W_{e,c,t} POP2064_{e,c,t}$$

$W_{e,c,t}$ corresponds to the education (e) weights specific to the average educational group of the country c at time t as calculated in the previous section. $POP2064_{e,c,t}$ is the population aged 20–64 from country c at time t with an educational attainment e . $SAF_{c,t}$ is the Skill-adjustment factor specific to country c at time t .

The population projection outcomes by education are sourced from the WIC (see previous section), in which the educational attainment variable is categorical rather than in number of years of schooling. Therefore, education-weights are calculated based on the assumed number of years of education for each level (see data source and selection of the sample). For instance, there were 76.4 M persons aged 20–64 with a postsecondary education (16 years of schooling) in USA in 2020 ($POP2064_{e6,USA,2020}$). Since the average number of years of schooling in 2020 was 13.2, the value of $W_{e6,USA,2020}$ is 2.45. Multiplying $W_{e6,USA,2020}$ by $POP2064_{e6,USA,2020}$ gives 187 M, which means, in other words, that these 76.4 M persons with upper secondary would have a human capital equivalent 187 M Americans who would have 9.3 years of schooling. The SAF value for USA in 2020 is 1.33. Adjusted for this factor, the human capital figure is 249 M. Assuming that within a country, the income is a good proxy for productivity, and that across countries, there is a linear relationship between the quality of education and productivity, these 76.4 M Americans with postsecondary education would have the productive capacity of 249 M persons with the 2015 world average's human capital. The HCWP for USA in 2020 is given by the sum of this calculation for each educational group.

Figure 3 shows the estimated human capital-weighted population together with the working-age (20–64) population size from 1970–2020 by continent and their projection up to 2100 according to the five SSP scenarios.

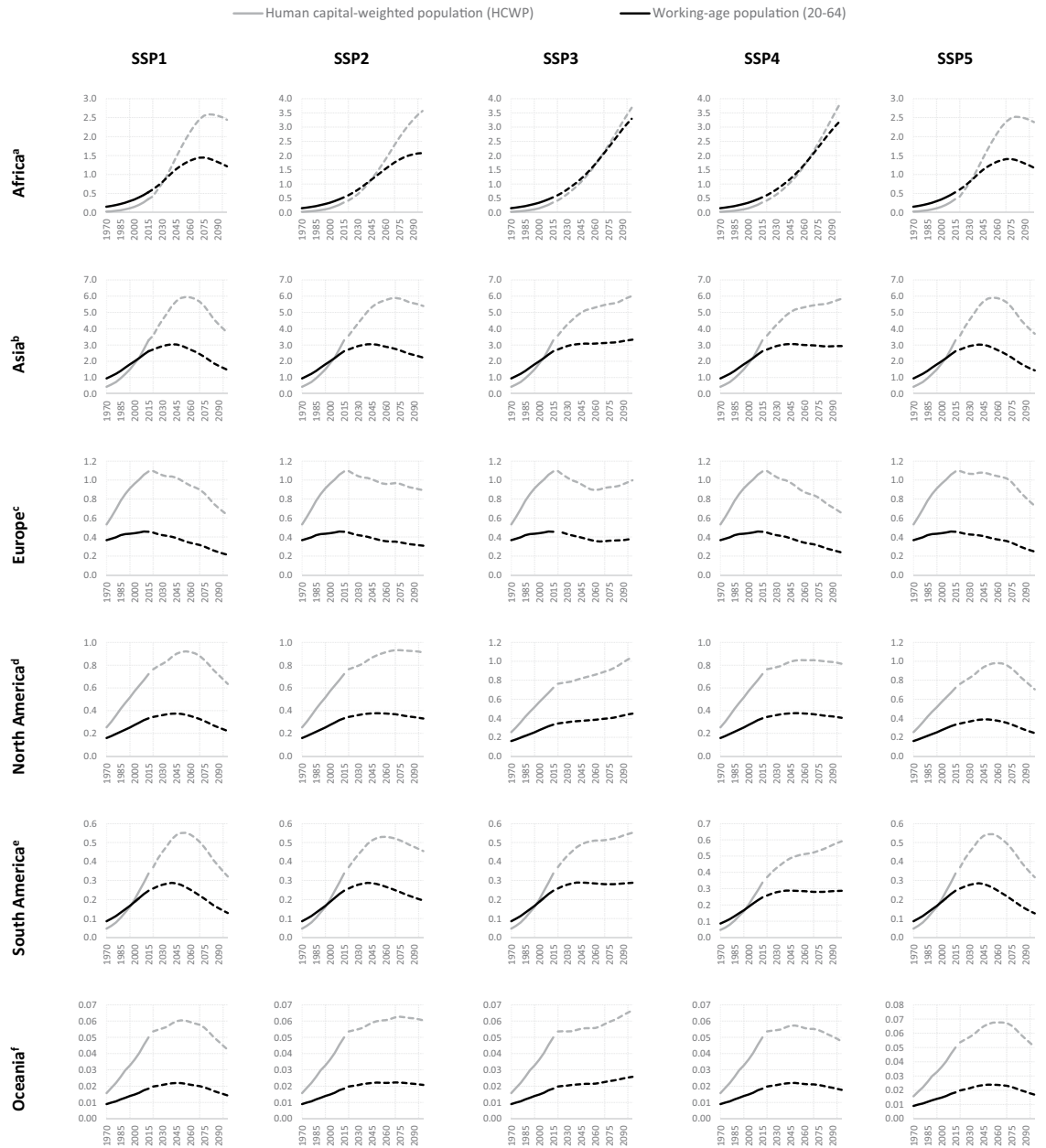
Data Records

The dataset OUTPUT_HCWP.csv available on Zenodo¹ contains three columns for analytical indicators: human capital-weighted population (HCWP), total working-age population aged 20–64 (WAP) and the total population (POP) for 185 countries (identified in “CountryName” and “CountryCode” columns). The “CountryCode” column corresponds to country codes from the Standard Country or Area Codes for Statistical Use (M49) provided by the Statistics Division of the United Nations Secretariat. “Region” and “Continent” correspond to the geographical classification of countries by the United Nations Statistics Division.

The total working-age population and the total population are taken directly from the Wittgenstein Centre's Data Explorer and are included in the dataset for information purposes, to give a better idea of the contribution of HCWP compared with this conventional indicator. Each variable contains 17,575 values, 185 countries X 10 periods from 1970–2015 + 5 scenarios X 17 periods from 2020 to 2100).

Technical Validation

We validate the HCWP data in the context of out-of-sample economic growth forecasts, by comparing the predictive content of the information of HCWP with that of human capital variables based exclusively on educational attainment. We estimate the econometric model proposed by Crespo Cuaresma⁵ using data for GDP per capita and capital stocks from the Penn World Table 10.1²⁹, population and educational attainment data from the WIC. The estimated models are based on an aggregate production function where technological progress is assumed to depend on innovation mechanisms (proxied alternatively using educational attainment and HCWP as the human capital stock variable) and technology adoption effects (included in the model as the interaction of the human capital variable and the level of income per capita). Using the estimated panel data models based on the global sample of countries and the period 1970–1995 in 5-year periods, out-of-sample predictions for all countries in 1995–2000 are obtained, and prediction error measures (root mean square error, mean absolute deviation and directional accuracy rate, that is, the share of correctly predicted directions of change) are computed. The exercise is repeated expanding the in-sample period to 1970–2000 and using the out-of-sample period 2000–2005, then for the in-sample period 1970–2005 and predictions for 2005–2010, and finally for



- a. Excluding British Indian Ocean Territory, Djibouti, Eritrea, French Southern Territories, Libya, Mauritania, Mayotte, Saint Helena, Seychelles, and Western Sahara
- b. Excluding Brunei Darussalam and Uzbekistan
- c. Excluding Åland Islands, Andorra, Faroe Islands, Gibraltar, Guernsey, Holy See, Isle of Man, Jersey, Liechtenstein, Monaco, San Marino, and Svalbard and Jan Mayen Islands
- d. Excluding Anguilla, Antigua and Barbuda, Barbados, Bermuda, Bonaire, Sint Eustatius and Saba, British Virgin Islands, Cayman Islands, Dominica, Greenland, Grenada, Montserrat, Saint Barthélemy, Saint Kitts and Nevis, Saint Martin (French Part), Saint Pierre and Miquelon, Sint Maarten (Dutch part), Turks and Caicos Islands, and United States Virgin Islands
- e. Excluding Bouvet Island, Falkland Islands (Malvinas), and South Georgia and the South Sandwich Islands
- f. Excluding American Samoa, Christmas Island, Cocos (Keeling) Islands, Cook Islands, Guam, Heard Island and McDonald Islands, Marshall Islands, Nauru, Niue, Norfolk Island, Northern Mariana Islands, Palau, Papua New Guinea, Pitcairn, Tokelau, Tuvalu, United States Minor Outlying Islands, and Wallis and Futuna Islands

Fig. 3 Working age population (WAP) and human capital-weighted population (HCWP), estimated (1970–2020) and projected (2020–2100) according to five SSP scenarios (in billions).

models estimated using data for 1970–2010 and forecasts for 2010–2015. We compare the average prediction errors for the different models over all of the out-of-sample observations in this design.

The results of the validation exercise for the models including HCWP as a predictor, the specifications with educational attainment and a benchmark model where total population is used as a human capital variable (without heterogeneity based on education or productivity) are presented in Table 3. In the global sample, the results for the models based on HCWP are relatively similar to those based on educational attainment, but the specifications that use HCWP as an explanatory variable excel in prediction for the subsample of low and

	Educational attainment	HCWP	Benchmark
Global Sample	0.235/0.174/0.744	0.236/0.174/0.733	0.245/0.178/0.737
High Income	0.196/0.152/0.653	0.186/0.143/0.653	0.196/0.152/0.638
Lower Middle Income	0.221/0.173/0.823	0.226/0.179/0.814	0.224/0.176/0.823
Upper Middle Income	0.278/0.198/0.772	0.286/0.206/0.73	0.301/0.209/0.763
Low Income	0.257/0.183/0.748	0.252/0.174/0.765	0.267/0.186/0.757

Table 3. Validation results based on predictive ability for GDP per capita growth. Each cell presents the root mean square error (RMSE), mean absolute deviation (MAD) and directional accuracy rate (DA), with bold figures marking the best performer (RMSE/MAD/DA).

high-income economies. Models based on educational attainment, on the other hand, dominate in terms of forecasting ability for middle-income economies.

While the HCWP model offers a more comprehensive approach by additionally considering educational quality, the complex interplay between human capital accumulation and economic growth at the global level may limit the predictive ability of econometric models built upon the new data. In particular, the assumption of parameter homogeneity which underlies our econometric specification may be restrictive for our global sample of countries, that contains economies with very different institutional, political, and cultural characteristics. Furthermore, synergies between quantity and quality of education are not necessarily linear or additive, and exploring such interactions could be a promising path of future research. Finally, the data used to construct the HCWP model may contain inherent variability and measurement issues, particularly when assessing educational quality across different global contexts. Future research could delve deeper into context-specific analyses to further explore these dynamics.

Usage Notes

The methodology used to produce the human capital-weighted population (HCWP) estimates has some limitations. First, human capital is assessed only in terms of the quality and quantity of education. This approach may oversimplify the multidimensional nature of human capital, as it fails to consider other critical dimensions such as health, which influences educational opportunities, labor force participation, and productivity³⁰. By excluding health, our analysis could potentially overestimate the HCWP and underestimate cross-country variations. Indeed, the HCWP takes into account only the gains in human capital resulting from changes in education, neglecting the potential impact of health variables among others. Future research iterations should explore the inclusion of health-related factors to capture a more comprehensive understanding of human capital.

Moreover, the education variable itself may oversimplify the diversity in educational attainment, potentially masking important nuances in educational backgrounds. This approach may not fully capture the varying impacts of different types of education on human capital accumulation. This is particularly true in countries where a large share of the population has post-secondary education, which includes very different specializations and skills.

Furthermore, the use of pooled censuses from different countries and years to estimate human capital weights may overlook country-specific nuances and variations in educational systems, labor markets, and economic conditions, which could significantly influence the relationship between education and human capital. It is also worth noting that although the countries included in our sample cover many degrees of development, many countries are missing from the analysis. Most censuses were from countries in the Americas, potentially limiting the generalizability of our findings to other regions with distinct socioeconomic and cultural contexts.

Additionally, our approach of adjusting education weights for quality using cross-sectional data from adult literacy assessments assumes uniform quality across different levels of education within a country. This assumption may overlook potential variations in educational quality that could impact human capital differently across various educational levels.

Finally, projecting future human capital-weights based on assumptions about educational trends and fertility rates introduces uncertainties related to demographic changes, economic developments, and policy interventions that may not materialize as projected. These uncertainties could potentially affect the accuracy of our long-term projections.

Code availability

Regression models to estimate of human capital-weights (Eqs. 1, 2 and 3) were run using the GLIMMIX procedure in SAS 9.4.

```
/*Model 1*/
proc glimmix data=data INITGLM;
title "Fixed all";
class sample SEXR;
model norm_sal=nbedu experience|experience SEXR / solution dist=gamma
link=log;
weight perwt;
where 0<norm_sal<15 and 20<=agegr<=64;
NLOPTIONS TECH=NRRIDG MAXITER=100;
run;
```



```

/*Model 2*/
proc glimmix data=data INITGLM;
title "Random intercept";
class sample SEXR;
model norm_sal=nbedu experience|experience SEXR / solution dist=gamma
link=log;
random intercept / subject=sample;
weight perwt;
where 0<norm_sal<15 and 20<=agegr<=64;
NLOPTIONS TECH=NRRIDG MAXITER=100;
run;

/*Model 3*/
proc glimmix data=data INITGLM;
title "Random intercept with interaction";
class sample SEXR;
model norm_sal=nbedu|nbedu_mean experience|experience SEXR / solution
dist=gamma link=log;
random intercept / subject=sample;
weight perwt;
where 0<norm_sal<15 and 20<=agegr<=64;
NLOPTIONS TECH=NRRIDG MAXITER=100;
run;

```

The code for the calculation of the Human capital weighted population estimates is available on Github (<https://github.com/gmarois/HCWP/blob/main/Code%20for%20HCWP.sas>).

Received: 29 January 2024; Accepted: 3 June 2024;

Published online: 12 June 2024

References

1. Marois, G., Crespo Cuaresma, J., Zellmann, J. G. & Reiter, C. Human Capital-weighted population estimates for 185 countries from 1970 to 2100 (2.0) *Zenodo*. <https://doi.org/10.5281/zenodo.11484066> (2024).
2. Bloom, D. E., Canning, D. & Sevilla, J. *The Demographic Dividend: A New Perspective on the Economic Consequences of Population Change*. (2003).
3. Golley, J. & Tyers, R. Demographic dividends, dependencies, and economic growth in China and India. *Asian Econ. Pap.* **11**, 1–26 (2012).
4. Mason, A., Lee, R. & Jiang, J. X. Demographic Dividends, Human Capital, and Saving. *J. Econ. Ageing* **7**, 106–122 (2016).
5. Crespo Cuaresma, J. Income projections for climate change research: A framework based on human capital dynamics. *Glob. Environ. Change* **42**, 226–236 (2017).
6. Crespo Cuaresma, J., Lutz, W. & Sanderson, W. C. Is the demographic dividend an education dividend? *Demography* **51**, 299–315 (2014).
7. Lutz, W., Crespo Cuaresma, J. & Sanderson, W. C. The demography of educational attainment and economic growth. *Science* **319**, 1047–1048 (2008).
8. Lutz, W. *et al.* Education rather than age structure brings demographic dividend. *Proc. Natl. Acad. Sci.* 201820362, <https://doi.org/10.1073/pnas.1820362116> (2019).
9. KC, S., Wurzer, M., Speringer, M. & Lutz, W. Future population and human capital in heterogeneous India. *Proc. Natl. Acad. Sci.* **115**, 8328 (2018).
10. Lutz, W. & KC, S. Global human capital: Integrating education and population. *Science* **333**, 587–592 (2011).
11. Martin, T. C. & Juarez, F. The impact of women's education on fertility in Latin America: Searching for explanations. *Int. Fam. Plan. Perspect.* **21**, 52–80 (1995).
12. Valkonen, T. Social inequalities in mortality. in *Demography: Analysis and Synthesis Volume 2: A Treatise in Population Studies* (eds. Caselli, G., Vallin, J. & Wunsch, G.) vol. 2 195–206 (Academic Press, London, UK, 2006).
13. Docquier, F. & Marfouk, A. International migration by education attainment, 1990–2000. in *International Migration, Remittances and the Brain Drain* (eds. Özden, C. & Schiff, M.) 38 (The World Bank, Washington, D.C., 2004).
14. Skirbekk, V. Fertility trends by social status. *Demogr. Res.* **18**, 145–180 (2008).
15. Kravdal, Ø. & Rindfuss, R. R. Changing Relationships between Education and Fertility: A Study of Women and Men Born 1940 to 1964. *Am. Sociol. Rev.* **73**, 854–873 (2008).
16. Marois, G., Sabourin, P. & Bélanger, A. How reducing differentials in education and labor force participation could lessen workforce decline in the EU-28. *Demogr. Res.* **41**, 125–160 (2019).
17. *Demographic and Human Capital Scenarios for the 21st Century*. (Publications Office of the European Union, Luxembourg, 2018).
18. Lutz, W. *et al.* Skills-adjusted human capital shows rising global gap. *Proc. Natl. Acad. Sci.* **118**, e2015826118 (2021).
19. Heyneman, S. P. International education quality. *Spec. Issue Honor Lewis C Solman* **23**, 441–452 (2004).
20. Marois, G., Bélanger, A. & Lutz, W. Population aging, migration, and productivity in Europe. *Proc. Natl. Acad. Sci.* **117**, 7690 (2020).
21. Marois, G., Gietel-Basten, S. & Lutz, W. China's low fertility may not hinder future prosperity. *Proc. Natl. Acad. Sci.* **118**, e2108900118 (2021).
22. Mincer, J. *Schooling, Experience, and Earnings*. (Columbia University Press, New York, 1974).
23. Psacharopoulos, G. & Patrinos, H. A. Returns to investment in education: a decennial review of the global literature. *Educ. Econ.* **26**, 445–458 (2018).
24. Van Biesebroeck, J. *How Tight Is the Link between Wages and Productivity? A Survey of the Literature*. 47 (2015).
25. KC, S. *et al.* Wittgenstein Center (WIC) Population and Human Capital Projections - 2023. *Zenodo*, <https://doi.org/10.5281/ZENODO.7921989> (2023).
26. Basten, S., Sobotka, T. & Zeman, K. Future fertility in low fertility countries. in *World Population and Human Capital in the 21st Century* (eds. Lutz, W., Butz, W. P. & KC, S.) 39–146 (Oxford University Press, Oxford, 2014).

27. Fuchs, R. & Goujon, A. Future fertility in high fertility countries. in *World Population and Human Capital in the 21st Century* (eds. Lutz, W., Butz, W. P. & KC, S.) 147–225 (Oxford University Press, Oxford, 2014).
28. Lutz, W., Butz, W. P. & KC, S. World Population & Human Capital in the Twenty-first Century: Executive Summary. <http://pure.iiasa.ac.at/11189/> (2014).
29. Feenstra, R. C., Inklaar, R. & Timmer, M. P. The Next Generation of the Penn World Table. *Am. Econ. Rev.* **105**, 3150–3182 (2015).
30. Gyimah-Brempong, K. & Wilson, M. Health human capital and economic growth in Sub-Saharan African and OECD countries. *Q. Rev. Econ. Finance* **44**, 296–320 (2004).

Acknowledgements

This work was supported by Research Fund for International Scientists of the National Natural Science Foundation of China [grant number 72250610226].

Author contributions

Conceptualization: G.M. Methodology: G.M., J.C.C. Investigation: G.M., J.C.C. Visualization: G.M. Data Collection: G.M., C.R. Validation: J.Z., J.C.C. Writing – original draft: G.M., J.C.C. Writing – review & editing: J.C.C., C.R.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to G.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024