International Institute for Applied Systems Analysis
IIASA www.iiasa.ac.at

Address: IIASA, Schlossplatz 1, A-2361 Laxenburg, Austria

Email: permissions@iiasa.ac.at

Working paper

Measuring Human Capital with Productivity-Weighted Labor Force: Methodology and Projections for China, India, the United States, and the European Union

Guillaume Marois (marois@iiasa.ac.at)
Stuart Gietel-Basten (sgb@ust.hk)
Jesus Crespo Cuaresma (jesus.crespo.cuaresma@wu.ac.at)
Jakob Gregor Zellmann (jakob.gregor.zellmann@s.wu.ac.at)
Claudia Reiter (reiter@iiasa.ac.at)
Wolfgang Lutz (lutz@iiasa.ac.at)
WP-24-005

Approved by:

Hans Joachim Schellnhuber
Director General
Date: 12 March 2024

Table of contents

ZVR 524808900

The authors gratefully acknowledge funding from IIASA and the National Member Organizations that support the institute (The Austrian Academy of Sciences; The Brazilian Federal Agency for Support and Evaluation of Graduate Education (CAPES); The National Natural Science Foundation of China (NSFC); The Academy of Scientific Research and Technology (ASRT), Egypt; The Finnish Committee for IIASA; The Association for the Advancement of IIASA, Germany; The Technology Information, Forecasting and Assessment Council (TIFAC), India; The Indonesian National Committee for IIASA; The Iran National Science Foundation (INSF); The Israel Committee for IIASA; The Japan Committee for IIASA; The National Research Foundation of Korea (NRF); The Mexican National Committee for IIASA; The Research Council of Norway (RCN); The Russian Academy of Sciences (RAS); Ministry of Education, Science, Research and Sport, Slovakia; The National Research Foundation (NRF), South Africa; The Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (FORMAS); The Ukrainian Academy of Sciences; The Research Councils of the UK; The National Academy of Sciences (NAS), USA; The Vietnam Academy of Science and Technology (VAST).

Abstract

This working paper provides a comprehensive overview of the methodology used to calculate a standardized and internationally comparable productivity-weighted labor force (PWLF) measure that takes into account both the education structure of the population and the quality of the educational system. Education-specific weights are calculated with a Mincerian earnings function on pooled data from all IPUMS-I censuses containing information on education, labor force status, and income. 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. To calculate the productivity-weighted labor force size, these adjusted weights are then applied to labor force estimates and projections. The analytical value of the PWLF is validated making use of prediction exercise for GDP growth applied to a panel dataset covering all countries of the world from 1970 to 2015 for which data are available. Finally, the paper provides a practical application by forecasting PWLF figures for China, India, the United States, and the European Union from 2020 to 2100. These forecasts are compared against other population indicators (total population size, working-age population, and labor force size), highlighting the importance of population heterogeneity in the analysis of demographic trends.

About the authors

Guillaume Marois is a research scholar at the International Institute for Applied Systems Analysis (Laxenburg, Austria) and a professor at the Asian Demographic Research Institute (ADRI) of Shanghai University (China). Contact: marois@iiasa.ac.at

Stuart Gietel-Basten is professor of Social Science and Public Policy at the Hong Kong University of Science and Technology. Contact: sgb@ust.hk

Jesus Crespo Cuaresma is professor of economics at the Vienna University of Economics and Business, and a research scholar at the International Institute of Advanced Systems Analysis (Laxenburg, Austria). Contact: jesus.crespo.cuaresma@wu.ac.at

Jakob Gregor Zellmann is a research scholar at the department of economics at the Vienna University of Economics and Business. Contact: jakob.gregor.zellmann@s.wu.ac.at

Claudia Reiter is a researcher at the Institute for Advanced Studies (Vienna, Austria) and a research scholar at the International Institute for Applied Systems Analysis (Laxenburg, Austria). Contact: reiter@iiasa.ac.at

Wolfgang Lutz is the Founding Director of the Wittgenstein Centre for Demography and Global Human Capital (IIASA, OeAW, University of Vienna) and Interim Deputy Director General for Science at the International Institute for Applied Systems Analysis. Contact: lutz@iiasa.ac.at

Acknowledgments

We would like to express our gratitude to Klaus Prettner for his support in the econometric modeling of productivity-weights.

Background

This working paper provides a comprehensive overview of the methodology used to calculate a standardized and internationally comparable productivity-weighted labor force (PWLF) measure that takes into account both the education structure of the population and the quality of the educational system. A number of recent papers have already analyzed the demographic challenges using a "productivity-weighted labor force", which multiplies the number of workers by a productivity factor based on their educational attainment (Marois, Bélanger, and Lutz 2020; Marois, Gietel-Basten, and Lutz 2021). However, the productivity-weights calculated in these studies are not suitable for international comparisons or time series for three main reasons. First, weights from a country- or region- specific data source (survey or census). The weights are thus relative to a reference category, which consists of a specific population group of the country surveyed (for example the "workers with X level of education of country Z"), and therefore cannot be generalized to broader international comparisons between countries with very different range of salaries.

Second, the approach assumes that the relative productivity gap between educational levels is constant over time. This is unlikely to be the case, since in a competitive market the relative advantage of a given level of education depends on many factors that change over time, such as the educational attainment of other workers and the availability of jobs requiring those skills (Psacharopoulos and Patrinos 2018).

Third, 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 world (Lutz et al. 2021). 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 underfunded (Heyneman 2004).

In this paper, we improve on the approach developed in previous studies to calculate productivity. Our methodology addresses existing limitations by constructing productivity-weights that allow consistent comparisons across countries and time periods. Education-specific weights are calculated with a Mincerian earnings function on pooled data from all IPUMS-I censuses containing information on education, labor force status, and income. 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 (Lutz et al. 2021). Country and time specific adjustment factors for education quality are derived from skills assessment surveys. To calculate the productivity-weighted labor force size, these adjusted weights are then applied to labor force estimates and projections. The analytical value of the PWLF is validated making use of prediction exercise for GDP growth applied to a panel dataset covering all countries of the world from 1970 to 2015 for which data are available. Finally, the paper provides a practical application by forecasting PWLF figures for China, India, the United States, and the European Union from 2020 to 2100. These forecasts are compared against other population indicators (total population size, working-age population, and labor force size), highlighting the importance of population heterogeneity in the analysis of demographic trends.

Methodology

Estimates of education-specific productivity-weight

Data source and selection of the sample

Our estimates of productivity-weights by educational attainment are based on an approach developed in other country-specific studies, using the wage as a proxy for productivity (Marois, Bélanger, and Lutz 2020; Marois, Gietel-Basten, and Lutz 2021). In order to derive generalizable weights linked to educational levels, we compiled all available censuses from the IPUMS-International database that included total personal income, labor force status and 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, respondents in the labor force with an income lower or equal to 0 were given a value of 1 for modeling purposes. Employed individuals with missing income data 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' income by dividing it 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 different ways, either the number of years of education or/and the educational attainment (with a varying number of categories). We chose to use the number of years of education as interest variable in the models, because broader categories can indeed hide a large heterogeneity. For instance, in a country where education is universal and mandatory until a certain age the lowest category would include all those who did not finish high school, while in developing countries, censuses usually distinguish those who did not attend any school from those who went to primary school and those who attended lower secondary level. For datasets only providing educational attainment in categories, we converted the categories 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.

We finally excluded from the sample those aged under 15 and over 75 and those outside of the labor force. 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.

Table 1. Selected census and their sample size

Country	Year	Total
Canada	1971	8,470
Canada	1981	24,620
Canada	1991	43,022
Canada	2001	47,594
Canada	2011	55,973
Colombia	1973	27,917
Dominican Republic	1981	7,746
Dominican Republic	2002	20,958
Mauritius	2000	25,538
Mexico	1995	27,841
Panama	1980	10,746
Panama	1990	11,967
Panama	2010	14,557
Puerto Rico	1990	23,558
Puerto Rico	2000	22,708
Puerto Rico	2005	12,338
Puerto Rico	2010	12,939
South Africa	1996	11,614
South Africa	2001	14,742
South Africa	2007	18,073
South Africa	2011	19,575
Trinidad and Tobago	1970	14,214
Trinidad and Tobago	2000	38,412
United States	1960	69,805
United States	1970	81,148
United States	1980	42,280
United States	1990	61,149
United States	2000	68,131
United States	2005	70,725
United States	2010	75,869
United States	2015	75,887
Total		1,060,116

Models

We estimated the productivity weights from Mincerian earnings function (Mincer 1974), which is widely used in the literature to measure the return on investment in education (Psacharopoulos and Patrinos 2018). We start by building a simple log-linear model with gamma distribution predicting the natural logarithm of the normalized income with the number of years of education, controlling for the experience and the sex, but without considering the design of the sample:

Eq. 1

$$Model\ 1$$
$$\ln(NOR\_INC) = \beta_0 + \beta_1 * NBEDU + \beta_2 EXPERIENCE \ + \beta_3 EXPERIENCE^2 + \beta_4 SEX + \varepsilon$$

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).
Eq. 2

$$Model\ 2$$
$$\ln(NOR\_INC) = \beta_{0j} + \beta_1 NBEDU + \beta_2 EXPERIENCE \ + \beta_3 EXPERIENCE^2 + \beta_4 SEX + \varepsilon_j$$

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 population (Psacharopoulos and Patrinos 2018). This model adds a parameter β_5 for the interaction between the number of years of education at the individual level and a country-level variable that refers to the average years of schooling for the population aged 25 to 54 (NBEDU_MEAN) in a respective census year.
Eq. 3

$$Model\ 3$$
$$\ln(NOR\_INC) = \beta_{0j} + \beta_1 * NBEDU + \beta_2 * EXPERIENCE \ + \beta_3 * EXPERIENCE^2 + \beta_4 * SEX + \beta_5 * NBEDU * NBEDU\_MEAN + \varepsilon_j$$

Parameters

Table 2 shows the parameter estimates for the three models employed. Despite income normalization, the covariance estimates in the model given by equation (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 equation (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 equation (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.

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

	Model 1	**Model 2**	**Model 3**
Random effect (group=census)		*Covariance*	*Covariance*
Intercept		0.097	0.023
		(0.040)	(0.009)
Fixed effect	*Estimate*	*Estimate*	*Estimate*
Intercept	-2.726	-2.796	-2.571
	(0.005)	(0.057)	(0.117)
NBEDU	0.119	0.139	0.237
	(0.000)	(0.000)	(0.003)
NBEDU_MEAN			-0.031
			(0.011)
NBEDU*NBEDU_MEAN			-0.009
			(0.000)
Experience	0.073	0.075	0.075
	(0.000)	(0.000)	(0.000)
Experience^2	-0.001	-0.001	-0.001
	(0.000)	(0.000)	(0.000)
SEX=Male	0.502	0.481	0.481
	(0.002)	(0.002)	(0.002)
N	1,060,116	1,060,116	1,060,116
Clusters		31	31
ICC		<0.001	<0.001

*All parameters are significant at $p<0.0001$

Using parameters from the model in equation (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

Eq 4. $$W_{e,c,t} = \frac{\exp(\beta_1 * NBEDU_{e,c,t} + \beta_5 * NBEDU * NB\_EDU\_MEAN_{c,t})}{\exp(\beta_1 * 9.3 + \beta_5 * 9.3 * NB\_EDU\_MEAN_{c,t})}$$

Figure 1 below presents weights by educational attainment and average years of schooling. In a country like the US in 2020 where the average number of years of education for the population aged 25-54 was 13.2 years, a worker with a postsecondary education would thus be weighted 2.26 times more than a worker with the world average number of years of education[1], and, assuming that income is a good proxy for productivity (Van Biesebroeck 2015), would thus be 2.26 times more productive.

[1] Since exp(0.237*16+-0.009*16*13.2)/exp(0.237*9.3+-0.009*9.3*13.2)=2.26

Figure 1. Productivity-weights ($W_{e,c}$) by educational levels and average years of schooling of the 25-54 in the country

1=world average level of education=9.3

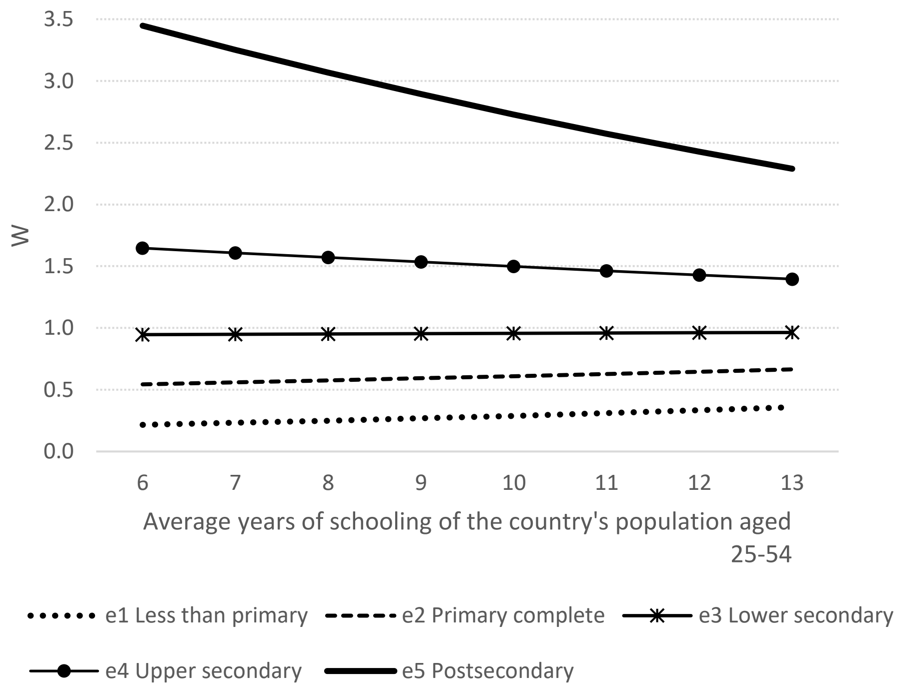
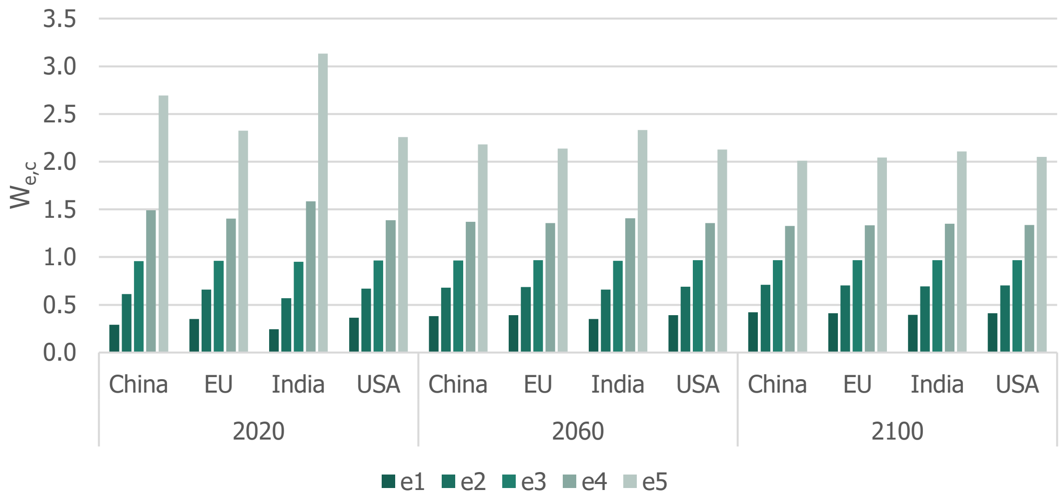


Figure 2 shows the projected values of the productivity-weights ($W_{e,c}$) at different points in time for the four regions, using the most recent update of the SSP2 (middle-of-the-road) scenario of the Wittgenstein Center Data Explorer for education forecasts (KC et al. 2024). Under this assumption of continuous progress with long-term global convergence of educational attainment (see Figure 4 in the section on population projections below), the productivity gap between educational levels gradually narrows and converges.

Figure 2. Projected values of productivity-weights ($W_{e,c}$)

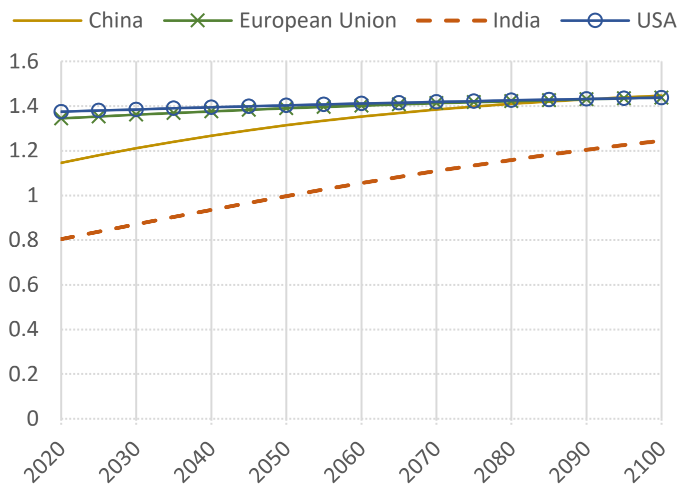


Factorizing for the quality of education

Lutz et al. (2021) 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 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 3 shows the resulting assumption from 2020 to 2100 on the skill-adjustment factor.

Figure 3. Assumption of the skill-adjustment factor (1=World average of 2015)



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 projection by age, sex, education and labor force participation

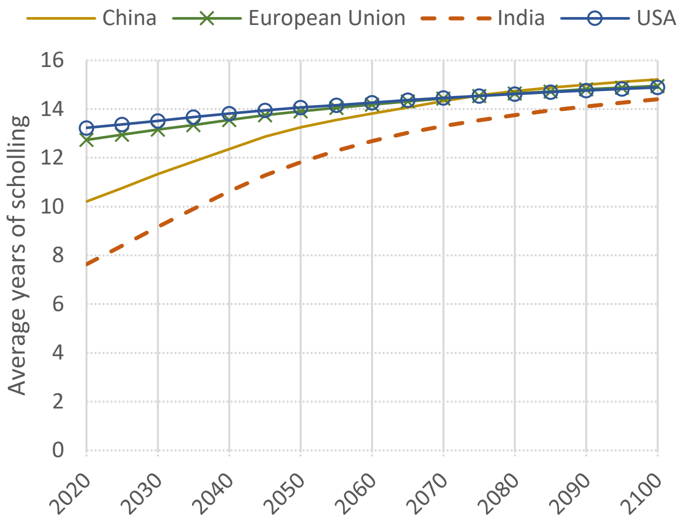
The calculation of the projected PWLF requires inputs from population projections by age, sex, and education. We used the most recent update of the SSP2 (middle-of-the-road) scenario of the Wittgenstein Center Data Explorer (KC et al. 2024), which includes the latest population estimates taking into account the impact of COVID-19 on the population structure as well as the latest estimates of demographic components as starting point. More details on the long-term assumptions can be found in Lutz et al. (2018; 2014).

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 average in- and out-migration rates of 1990-2020 until 2060, with convergence to net-zero total flows by 2100.

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, productivity weights are calculated based on the normal number of years of education for each level (see data source and sample selection). In Figure 4 we show the average years of schooling of the population aged 25-54, which is used in Eq. 4 in the calculation of the productivity weights.

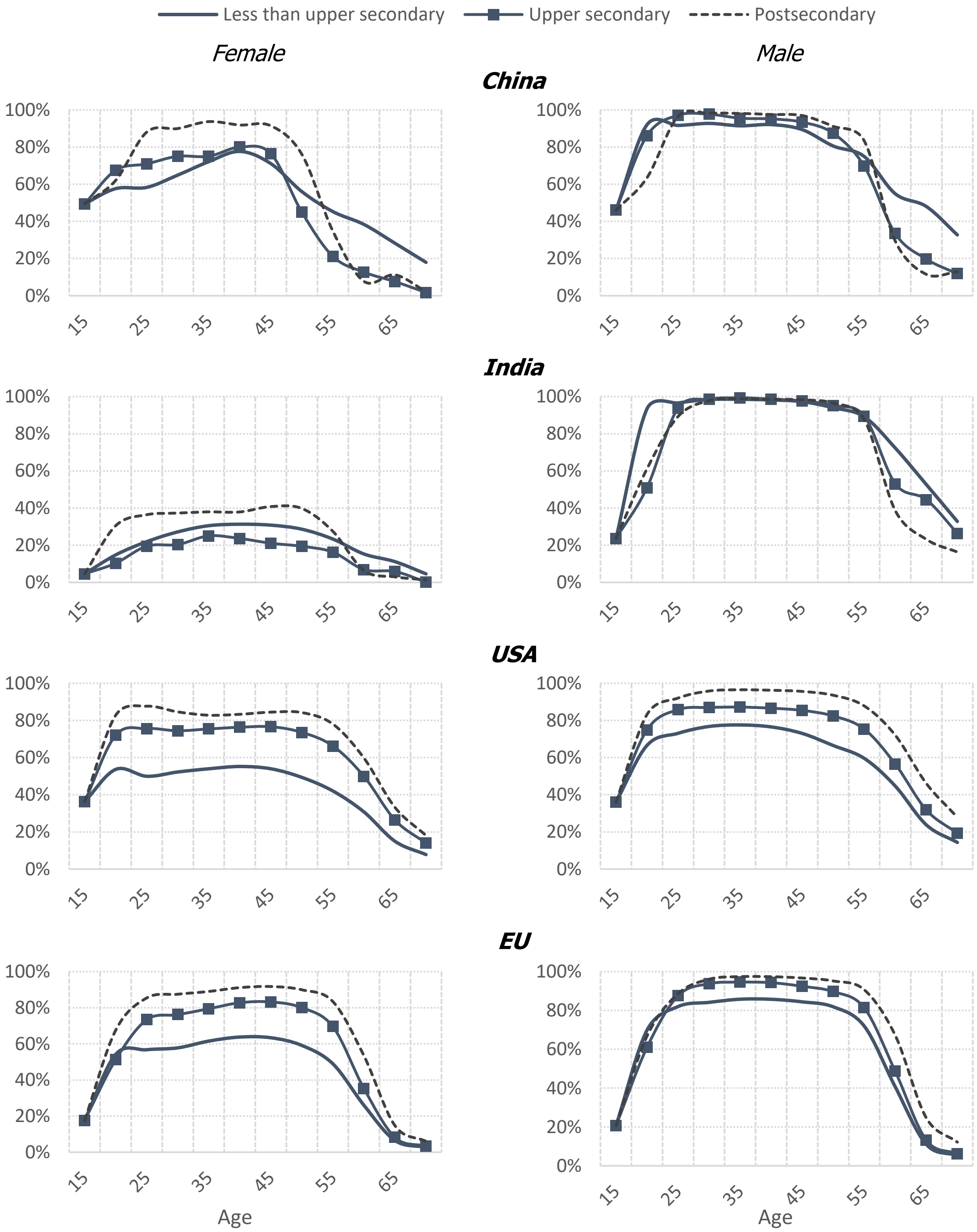
Figure 4. Projected average years of schooling of the population aged 25-54, 2020-2100



The future age- and education- specific fertility rates are determined on the basis of the combination of a large expert survey in the field of fertility studies (Basten, Sobotka, and Zeman 2014; Fuchs and Goujon 2014) and a model of historical analogy. The model assumed a trend towards global convergence in the very long term, resulting in slightly rising fertility in the European countries, also due to the ongoing process of fertility postponement; a continuation of the declining trend in India until the completion of the demographic transition; a slow recovery in China from the historically low level of recent years; and a slow decline in the USA, which has consistently had a higher fertility level than other Western countries for most of the last decades. By 2100, total fertility rates are projected to be 1.5 in China, 1.7 in the European Union, 1.6 in India and 1.7 in the United States.

To estimate the labor force size, we follow the state-of-the-art approach of superimposing group-specific participation rates at t on the population outputs (Marois, Gietel-Basten, and Lutz 2021; Van Hook et al. 2020; Marois, Bélanger, and Lutz 2020; Loichinger 2015; European Commission 2015). In our case, the projection started in 2020 with country-specific labor force participation rates by age, sex, and education calculated from the China General Social Survey 2010-2017, the Periodic Labour Force Survey (2017/18) of India, the European Labor Force Survey 2014-2019 and the American Census 2015 sample from IPUMS-I. Figure 5 shows these estimates.

Figure 5. Assumptions on labor force participation rates by age, sex, and education, China, India, United State (USA), and the European Union (EU)



We then developed two scenarios. The first one assumes constant group-specific rates, as do most other labor force projections (Van Hook et al. 2020; European Commission 2015; Loichinger 2010). Although group-specific rates are assumed to remain constant over the projection period, rates at the aggregate level change due to the compositional effects (such as increasing educational attainment among younger cohorts of women).

The second scenario assumes a convergence by 2100 of group-specific rates towards those currently observed in the USA. This scenario mainly affects the labor force participation rates of women for all groups in India, as this region is currently more affected by gender disparity than elsewhere (Batra and Reio 2016) and has among the lowest female labor force participation rates in the world (International Labour Organization 2022). Female labor force participation tends to follow a U-shaped relationship with economic development (Olivetti 2013; Tam 2011), and accordingly, this scenario assumes that India will also follow this pattern over the century. To a lesser extent, this convergence scenario also affects the participation rate of women in China aged 50 and over, as current regulations on the pension system still encourage early retirement. As concerns about the ageing population increase and the economy modernises with less physically demanding jobs, there are already some discussions to relax these regulations (Feng et al. 2019).

Calculation of the productivity weighted labor force size

For a country c at time t, the productivity-weighted labor force size (PWLF) is calculated as follows:

$$PWLF_{c,t} = SAF_{c,t} \sum_{e=1}^{k} W_{e,c,t} LABOR_{e,c,t}$$

$W_{e,c,t}$ corresponds to the productivity-weights by education level e specific to the average educational group of the country c at time t as calculated in the previous section. $LABOR_{e,c,t}$ is the population in the labor force. $SAF_{c,t}$ is the Skill-adjustment factor specific to country c at time t.

For instance, there were 70M workers with a postsecondary education in USA in 2020 ($LABOR_{e6,USA,2020}$), which corresponds to 16 years of schooling. As the average number of years of schooling in 2020 was 13.2, the value of $W_{e6,USA,2020}$ is 2.26. When multiplying $W_{e6,USA,2020}$ by $LABOR_{e6,USA,2020}$, we obtain 158M, which means, in other words, that these 70M persons with postsecondary education would have a productivity equivalent to 158M Americans who would have 9.3 years of schooling. The SAF value for USA in 2020 in then 1.37. When adjusting for this factor, we then have 216M, which would mean that after adjusting for their skills and productivity, these 70M American with postsecondary education are equivalent to 216M workers having the world average's productivity and skills. The PWLF for USA in 2020 is then the sum of this calculation for each education group.

Validation

To validate the analytical value of the PWLF, we used this indicator in the econometric model proposed by Crespo Cuaresma (2017) to predict 10-year GDP growth in a global sample of countries. We first constructed a panel dataset of 297 cross-sectional observations of countries/time periods of PWLF from 1970 to 2015 obtained by merging all IPUMS-I censuses ("Integrated Public Use Microdata Series, International: Version 7.3 [Dataset]"

2020) that provide labor force participation rates by age, sex, and education with their closest in time estimates from the Wittgenstein Data Explorer (WIC 2019). We then supplemented this dataset with GDP and capital stocks data from Penn World Table 10.1 (Feenstra, Inklaar, and Timmer 2015). From this cross-sectional data, 145 observations of 10-year changes are paired.

We performed a cross validation exercise where the models are trained on a randomly chosen subsample containing 70% of the data to predict the remaining 30%. Pseudo out of sample measures of fit are then calculated (root mean squared error, RMSE, mean absolute error, MAE, and R-squared) and the process is repeated, 10,000 times. Subsequently, the means of the measures of fit of the 10,000 iterations are calculated.

The models under consideration are aimed to explain GDP growth rate for 10-year periods and differ by the population related variables, either the population size (p_POP), the working-age population size (p_WA), the labor force size (p_LF) and the PWLF size (p_PWLF). Models are estimated with and without the inclusion of time and country specific fixed effects. Each of them controls for the initial GDP of the respective period and the capital growth in addition to the population-based covariates.

The results are presented in Table 3, which shows that p_PWLF models yield the lowest root mean square error (RMSE) and mean absolute error (MAE) values, while having higher R2, suggesting a higher superior predictive ability for economic growth compared to the other population variables. The improved performance highlights potential informative value of the PWLF as a predictor of the productive capacity of a country.

Table 3. Validation results based on predictive ability for 10-year GDP growth. *Root-mean-squared-errors (RMSE) and mean absolute errors (MAE) are normalized to 1 for the model using the PWLF. Bold figures mark the best performer.*

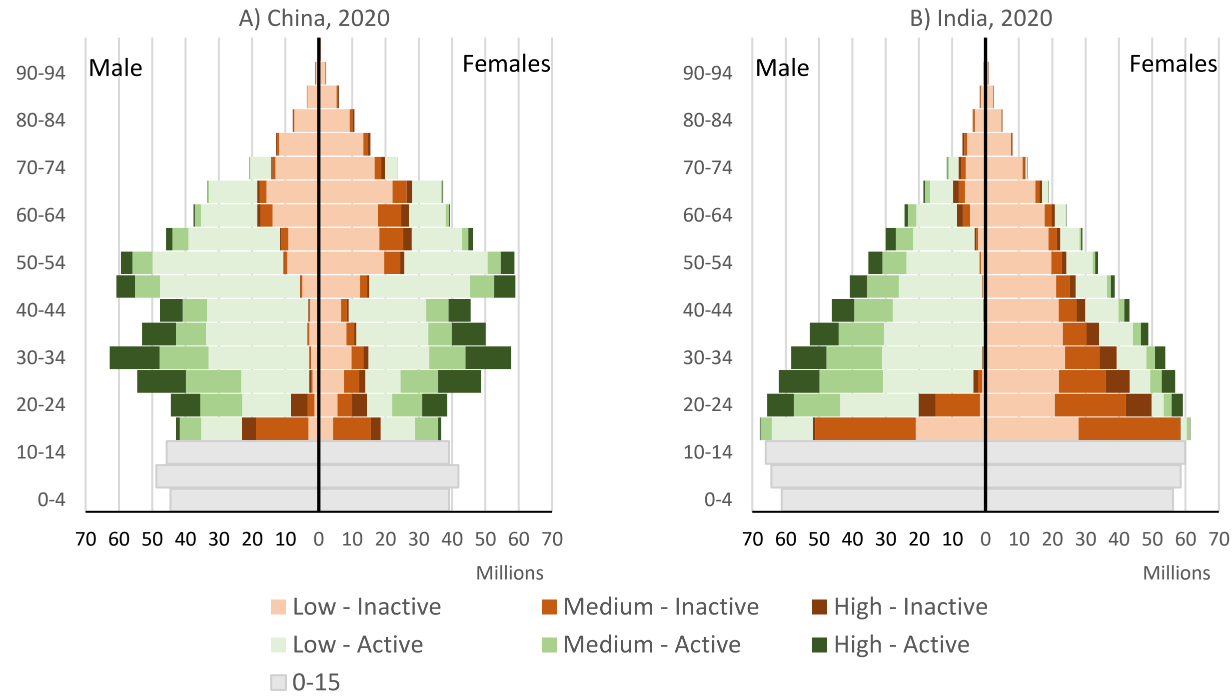
	Measure	**Model**			
		p_WA	*p_LF*	*p_PWLF*	*p_POP*
With fixed effects for country and time period	**RMSE**	1.053	1.026	**1.000**	**1.000**
	MAE	1.013	1.016	**1.000**	1.008
	R2	0.697	0.685	**0.711**	0.680
	Res. Std Error	0.191	0.195	**0.186**	0.195
Without fixed effects for country and time period	**RMSE**	1.020	1.028	**1.000**	1.014
	MAE	1.034	1.028	**1.000**	1.010
	R2	0.429	0.404	**0.516**	0.378
	Res. Std Error	0.204	0.209	**0.188**	0.213

1. Each model controls for the initial GDP and the capital growth.

Projections of the PWLF for China, India, the United States, and the European Union

Figure 6 showing the age-, sex-, education, and labor force pyramids for China and India and 2020 already highlights some major structural changes in the population composition that can be expected to have long lasting consequences. The Indian population is much younger and has a much higher proportion with no or only basic education, particularly among women. While over the past decades China heavily invested in universal primary and secondary education of all girls and boys, India until recently left a significant proportion its population behind. The much lower education of women also led to a slower fertility decline and thus a younger and still growing population. Female labor force participation in India is also extremely low, among the lowest in the world. This is true at all ages and levels of education, despite a fertility rate that is now below replacement level. Given the central importance of human capital for economic growth this also contributed to the much better economic performance of China over the past decades.

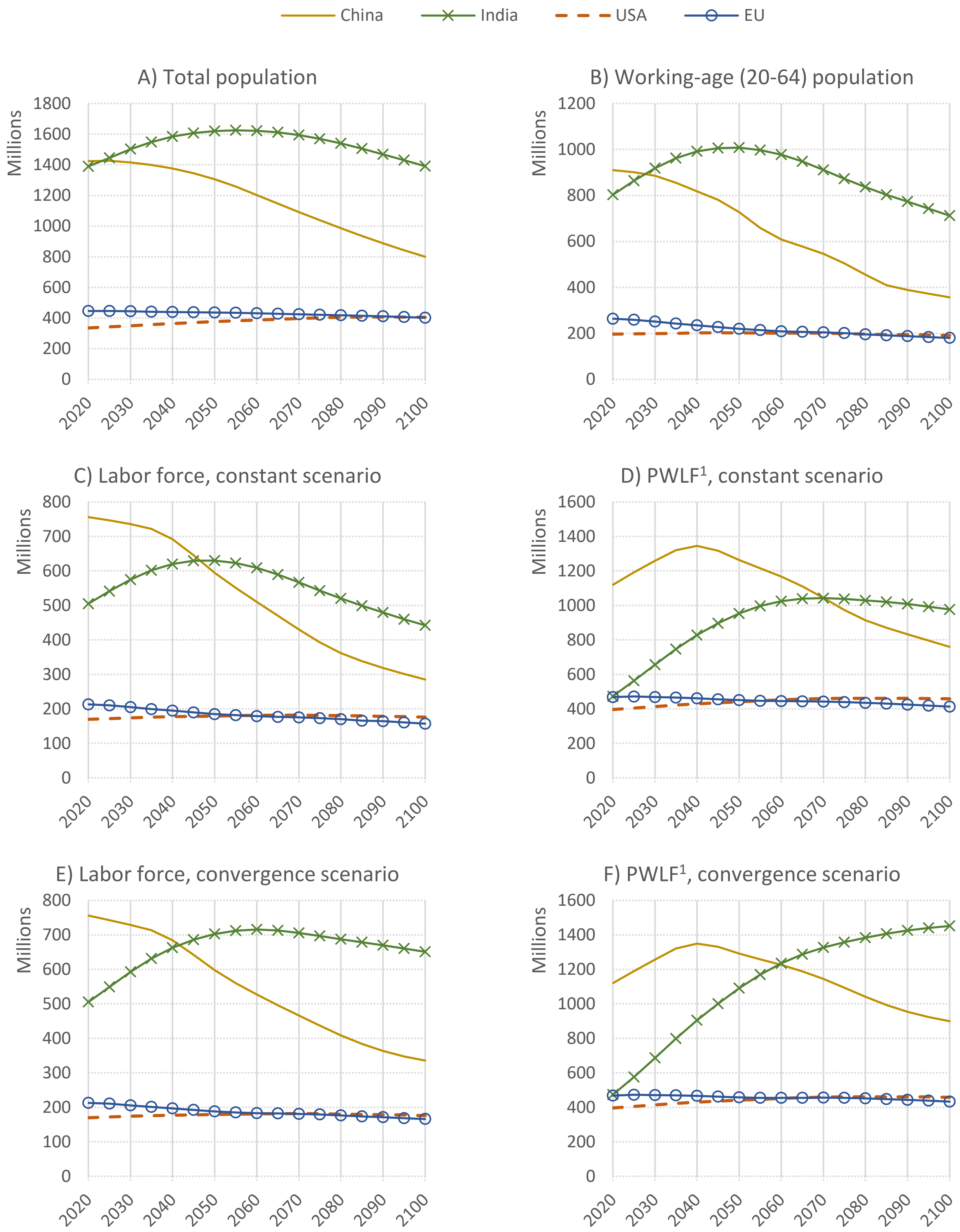
Figure 6A and 6B: Education and age pyramids for China (left) and India (right) for 2020



Source: Authors' calculations based on WIC 2023, Periodic Labor Force Survey, and China General Social Survey.

What do these differences in demographic structures imply for the future? As Figure 7 illustrates, the timing when India is likely to catch up with China demographically and economically greatly depends on the indicators used. While India already surpassed China in terms of total population size (Fig 7A), in terms of the size of the so-called "working age population" aged 20-64 (a misnomer because not everybody in this age range works and many work outside the range) India will catch up before 2030 as shown in Fig. 7B. If we consider actual labor force participation rates by age, sex, and education instead of just age groups and keep them constant over time, the picture looks quite different with India only catching up in the late 2040s (Fig 7C). If we finally take the much higher educational attainment of the Chinese population into account and assume that education is a proxy of productivity, then in terms of the size of the productivity weighted labor force China will still be more than two times stronger than India for the coming two decades and India will only catch up in 2070 (Fig 7D). Since this indicator is the most relevant for economic strength, this implies that for most of the next half century China will clearly remain the stronger economic power, despite of its low fertility and rapid population ageing.

Figure 7. Projected population outcomes, China, India, United State (USA) and the European Union (EU), 2020-2100.



1. The productivity-weighted labor force (PWLF) corresponds to the number of workers with average global human capital that would be equivalent in productive capacity to a country's actual labor force

Figure 7 also compares India and China with the USA and the EU. Two patterns immediately emerge. The first is how these four global demographic powerhouses compare to each other in terms of the productivity weighted labor force over the coming decades. The gap between China and USA/EU will widen until 2040, as China's human capital continues to expand despite of low birth rates. India's productivity weighted labor force – which is roughly at the same level as in the USA or EU today, despite of a much larger number of people – has the potential to rise to be more than twice the current level by mid-century. The second is the nature of the relative trajectories going into the longer-term future. China will need to rapidly adapt its economy in order to release the full potential of its new demographic structures. India, meanwhile, will not only need to promote the right conditions for productivity to increase (through gender equality, education and so on) but also turn this into growth through growing the labor force and investing in infrastructure and fixed capital. In comparison, the story of the EU and the USA look more 'stable' with their fraction of the world's human capital gradually declining, a trend also driven by rising human capital in Africa. The EU and the USA also can benefit from a flexible approach to immigration which, as other studies have shown, can also boost productivity and offset perceived demographic challenges, if managed in the right way (Marois, Bélanger, and Lutz 2020). This slower, perhaps more predictable pace of demographic change over the coming decades could work in favor of 'Old Europe' and the USA as they attempt to deal with economic and social issues which are already well in train (Coleman and Basten 2015).

While India has surpassed China as the world's most populous country, China is likely to remain the dominant economic power for the coming decades due to more favorable demographic structures. Specifically, China's higher levels of education and human capital, higher labor force participation rates, and a still sizable working-age population mean China will maintain a much larger productivity-weighted labor force over the next 50 years.

India's continuing population growth and enormous youth population could eventually become an economic advantage, but only if India makes substantial investments in education and effort in reducing gender inequality in the labor force. Boosting female education and labor force participation will be critical in determining when India can catch up to China economically.

For now, demography favors China remaining the world's largest economy for the foreseeable future. Policymakers in both nations must account for their shifting demographic structures when charting their countries' economic futures. China's debate on raising the female retirement age from 55 could significantly expand the female labor force for middle-aged and older women. However, recent policy shifts encouraging women to leave work to have babies may undermine female participation.

Limitations

The methodology used to estimate productivity weights based on education-specific indicators has several limitations that should be considered. The assumption of using income as a proxy for productivity might oversimplify the complex relationship between education, skills, and actual productivity levels. The productivity-weighted labor force (PWLF) indeed takes into account only the gain in productivity resulting from changes in education. Increases in productivity resulting from progress in technologies or improvements in institutional organization are not considered.

Furthermore, when the number of years of schooling was not available in censuses, they are estimated from the educational attainment, assuming the same number of years of schooling for a same educational attainment. However, there is heterogeneity across countries in the actual number of years corresponding to a given educational level (Potančoková, KC, and Goujon 2014). The education variable can also oversimplify the diversity in educational attainment, potentially masking important nuances in educational backgrounds. This approach may not capture the full impact of different types of education on productivity.

The use of pooled censuses from different countries and years to estimate productivity weights may overlook country-specific nuances and variations in educational systems, labor markets, and economic conditions. Additionally, although the countries included cover many degrees of development, many world countries are missing in the sample used to estimate the productivity weights. Most censuses were from countries in the Americas, potentially limiting the generalizability of the findings to other regions. This approach might not fully capture the unique factors influencing income differences at a national level. Moreover, adjusting education-weights for quality using cross-sectional data from adult literacy assessments assumes uniform quality across different levels of education within a country and overlooks potential variations in educational quality that could impact productivity differently.

Finally, projecting future productivity-weights based on assumptions about educational trends, fertility rates, labor force participation rates, and convergence towards specific scenarios introduces uncertainties related to demographic changes, economic developments, and policy interventions that may not materialize as projected.

These limitations highlight the need for caution when interpreting the estimated productivity-weights based on education-specific indicators and emphasize the importance of considering these constraints in drawing conclusions about the relationship between education and productivity.

Conclusion

This working paper presented a novel methodology to calculate a standardized and internationally comparable productivity-weighted labor force measure that accounts for both the educational attainment of the population and the quality of the educational system. The indicator is shown to have superior predictive power for GDP growth compared to conventional population indicators like total population, working-age population, or labor force size.

The paper provided practical illustrations by forecasting the productivity-weighted labor force for key regions like China, India, the United States, and the European Union until 2100, highlighting heterogeneity in demographic trends. Overall, the productivity-weighted labor force measure offers a more nuanced perspective on the productive capacity of national workforces, informing policy debates and analysis related to economic growth, human capital investment, and the socioeconomic implications of population aging and educational expansion. Future research could further refine the methodology and apply it to a broader range of countries and contexts.

References

Basten, Stuart, Tomáš Sobotka, and Kryštof Zeman. 2014. "Future Fertility in Low Fertility Countries." In *World Population and Human Capital in the 21st Century*, edited by Wolfgang Lutz, William P. Butz, and Samir KC, 39–146. Oxford: Oxford University Press. http://ukcatalogue.oup.com/product/9780198703167.do.

Batra, Renu, and Thomas G. Reio. 2016. "Gender Inequality Issues in India." *Advances in Developing Human Resources* 18 (1): 88–101. https://doi.org/10.1177/1523422316630651.

Coleman, David, and Stuart Basten. 2015. "The Death of the West: An Alternative View." *Population Studies* 69 (sup1): S107–18. https://doi.org/10.1080/00324728.2014.970401.

Crespo Cuaresma, Jesús. 2017. "Income Projections for Climate Change Research: A Framework Based on Human Capital Dynamics." *Global Environmental Change* 42: 226–36. https://doi.org/10.1016/j.gloenvcha.2015.02.012.

European Commission. 2015. "The 2015 Ageing Report: Economic and Budgetary Projections for the 28 EU Member States (2013-2060). European Economy 3/25." Brussels, Belgium: European Commission. http://ec.europa.eu/economy_finance/publications/european_economy/2015/pdf/ee3_en.pdf.

Feenstra, Robert C., Robert Inklaar, and Marcel P. Timmer. 2015. "The Next Generation of the Penn World Table." *American Economic Review* 105 (10): 3150–82. https://doi.org/10.1257/aer.20130954.

Feng, Qiushi, Wei-Jun Jean Yeung, Zhenglian Wang, and Yi Zeng. 2019. "Age of Retirement and Human Capital in an Aging China, 2015–2050." *European Journal of Population* 35 (1): 29–62. https://doi.org/10.1007/s10680-018-9467-3.

Fuchs, Regina, and Anne Goujon. 2014. "Future Fertility in High Fertility Countries." In *World Population and Human Capital in the 21st Century*, edited by Wolfgang Lutz, William P. Butz, and Samir KC, 147–225. Oxford: Oxford University Press. http://ukcatalogue.oup.com/product/9780198703167.do.

Heyneman, Stephen P. 2004. "International Education Quality." *Special Issue In Honor of Lewis C. Solman* 23 (4): 441–52. https://doi.org/10.1016/j.econedurev.2003.10.002.

"Integrated Public Use Microdata Series, International: Version 7.3 [Dataset]." 2020. Minnesota Population Center. https://international.ipums.org/.

International Labour Organization. 2022. "Labor Force Participation Rate, Female (% of Female Population Ages 15-64). ILOSTAT Database."

KC, S., M. Dhakad, M. Potančoková, S. Adhikari, D. Yildiz, M. Mamolo, T. Sobotka, et al. 2024. "Updating the Shared Socioeconomic Pathways (SSPs) Global Population and Human Capital Projections." IIASA Working Paper WP-24-003. Laxenburg, Austria.

Loichinger, Elke. 2010. "Der Fehlende Faktor: Zur Bedeutung Altersspezifischer Arbeitszeit Für Das Erwerbspotenzial." In *Wirtschaftspolitische Herausforderungen Des Demografischen Wandels*, 35–55. Springer.

———. 2015. “Labor Force Projections up to 2053 for 26 EU Countries, by Age, Sex, and Highest Level of Educational Attainment.” *Demographic Research* 32 (15): 443–86. https://doi.org/10.4054/DemRes.2015.32.15.

Lutz, Wolfgang, W. P. Butz, and Samir KC. 2014. “World Population & Human Capital in the Twenty-First Century: Executive Summary.” Other. October 2014. http://pure.iiasa.ac.at/11189/.

Lutz, Wolfgang, Anne Goujon, Samir KC, Marcin Stonawski, and Nikolaos Stilianakis, eds. 2018. *Demographic and Human Capital Scenarios for the 21st Century*. Luxembourg: Publications Office of the European Union.

Lutz, Wolfgang, Claudia Reiter, Caner Özdemir, Dilek Yildiz, Raquel Guimaraes, and Anne Goujon. 2021. “Skills-Adjusted Human Capital Shows Rising Global Gap.” *Proceedings of the National Academy of Sciences* 118 (7): e2015826118. https://doi.org/10.1073/pnas.2015826118.

Marois, Guillaume, Alain Bélanger, and Wolfgang Lutz. 2020. “Population Aging, Migration, and Productivity in Europe.” *Proceedings of the National Academy of Sciences* 117 (14): 7690. https://doi.org/10.1073/pnas.1918988117.

Marois, Guillaume, Stuart Gietel-Basten, and Wolfgang Lutz. 2021. “China’s Low Fertility May Not Hinder Future Prosperity.” *Proceedings of the National Academy of Sciences* 118 (40): e2108900118. https://doi.org/10.1073/pnas.2108900118.

Mincer, Jacob. 1974. *Schooling, Experience, and Earnings*. Human Behavior & Social Institutions, No. 2. New York: Columbia University Press. http://papers.nber.org/books/minc74-1.

Olivetti, Claudia. 2013. “The Female Labor Force and Long-Run Development.” National Bureau of Economic Research Working Paper Series No. 19131.

Potančoková, Michaela, Samir KC, and Anne Goujon. 2014. “Global Estimates of Mean Years of Schooling: A New Methodology.” Interim Report IR-14-005. Laxenburg, Austria: International Institute for Applied Systems Analysis (IIASA). http://www.iiasa.ac.at/publication/more_IR-14-005.php.

Psacharopoulos, George, and Harry Anthony Patrinos. 2018. “Returns to Investment in Education: A Decennial Review of the Global Literature.” *Education Economics* 26 (5): 445–58. https://doi.org/10.1080/09645292.2018.1484426.

Tam, Henry. 2011. “U-Shaped Female Labor Participation with Economic Development: Some Panel Data Evidence.” *Economics Letters* 110 (2): 140–42. https://doi.org/10.1016/j.econlet.2010.11.003.

Van Biesebroeck, Johannes. 2015. “How Tight Is the Link between Wages and Productivity? A Survey of the Literature.” Conditions of Work and Employment Series 54. Geneva: International Labour Office.

Van Hook, Jennifer, Alain Bélanger, Patrick Sabourin, and Anne Morse. 2020. “Immigration Selection and the Educational Composition of the US Labor Force.” *Population and Development Review* 46 (2): 321–46. https://doi.org/10.1111/padr.12315.

WIC. 2019. “Wittgenstein Centre Data Explorer Version 2.0.” 2019. www.wittgensteincentre.org/dataexplorer.