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

The Geography of Future Human Capital in Europe: Regional Dynamics of Education and Aging

[Andrea Tamburini, tamburini@iiasa.ac.at]

WP-26-004

Approved by:

Name: Anne Goujon
Program: POPJUS
Date: 01 July 2026

Table of contents

ZVR 524808900

The authors gratefully acknowledge funding from the European Union under grant agreement No 101081369 (SPARCCLE).

Abstract

The study of human capital formation and accumulation at the subnational level, including in the European context, has remained largely confined to historical and descriptive analyses rather than forward-looking projections. Spatially explicit projections of human capital, understood as the joint product of demographic structure (age and sex composition) and educational attainment, thus remain largely unavailable. This paper addresses this gap by developing a Bayesian time series model to project educational attainment across European NUTS-2 regions under the five Shared Socioeconomic Pathways (SSPs). The model works in log-return space, leverages the compositional nature of educational attainment distributions, and combines Bayesian uncertainty quantification with Iterative Proportional Fitting to ensure consistency with national SSP projections and regional demographic structures. The resulting dataset covers 215 NUTS-2 regions across 21 European countries, including three adult age groups from age 25 onwards and three education categories, spanning 2025 to 2100. The projections show a widespread yet heterogeneous increase in tertiary educational attainment, confirming convergence trends noted in existing literature, while the model also reveals that educational expansion unfolds at markedly different paces across regions, including within the same country, and a consistent female advantage across age groups and time horizons. Across European regions, the interplay between educational expansion and population ageing gives rise to diversified dynamics: rising attainment rates can coexist with a shrinking number of highly educated young adults where younger cohorts decline, while the gradual entry of more educated cohorts into older age groups reshapes the size and educational profile of the elderly population, with these processes often unfolding quite differently across regions within the same country.

About the authors

Andrea Tamburini is a researcher at the Population and Just Societies (POPJUS) Program in IIASA and PhD candidate at the Department of Demography at the University of Vienna (Contact: tamburini@iiasa.ac.at).

Acknowledgments

The authors gratefully acknowledge funding from the European Union under grant agreement No 101081369 (SPARCCLE).

1 Introduction

Human capital, the stock of knowledge, skills, and capabilities embodied by the population, plays a central role in shaping long-term economic and societal outcomes (Lutz 2009, 2017). Understanding its formation and accumulation dynamics is essential for understanding economic growth (Romer 1989; Galor and Tsiddon 1997; Pelinescu 2015; Zhu and Li 2017; X. Yang 2020) and resilience (Marois et al. 2019; Égert et al. 2020; Castelló-Climent and Domenech 2022). Beyond the economic dimension, human capital is equally relevant for climate adaptation (Reid et al. 2009; Striessnig et al. 2013; Sestito et al. 2025) and for national and regional strategies targeting sustainable development (Duran et al. 2015).

Yet the capacity to anticipate how human capital stocks will evolve, i.e., to produce reliable human capital projections, depends, at its most fundamental level, on understanding the demographic processes through which they are formed and renewed. Unlike age structure alone, which is ultimately a mechanical property of population composition, human capital is both produced through education systems and deployed through labour markets in ways that directly shape productivity, innovation, and long-run competitiveness.

The methodological point of departure for human capital projections can be found in Lutz and Goujon (2001) who demonstrated that multi-state cohort-component population projections could be extended to incorporate educational attainment as a third dimension of population structure alongside age and sex. This seemingly technical innovation carried profound substantive implications: since fertility, mortality, and migration all strongly vary by level of educational attainment, ignoring the educational composition of the population introduces systematic bias into long-range demographic projections. Building on this foundation, Lutz et al. (2008) were among the first to disentangle the separate effects of age structure and educational attainment on economic growth using cross-national panel data, finding that education matters on top of age structure. This directly challenged the existing literature, suggesting that the working-age share of the population was the primary demographic engine of economic change, and implied that the so-called "demographic dividend" is, to some extent, an education dividend (Lutz et al. 2019). This view has since been widely discussed and criticized (e.g., by Kotschy et al. 2020), but the fact that education in one essential aspect of human capital formation remains unchallenged.

The key mechanism through which education reshapes populations and economies is what Lutz (2013) termed "demographic metabolism", following earlier work by Ryder (1965): as younger and better-educated cohorts gradually replace older ones through the succession of cohorts, the educational composition of the workforce and the population more broadly, changes in predictable, cohort-traceable ways. Slowly, over decades, but with a high degree of foreseeability, this mechanism makes long-range projection both tractable and policy-relevant (Lutz 2013; Lutz et al. 2019). From this perspective, education is not merely a correlate of development but one of its root causes.

The empirical and methodological architecture integrating education into population projections at a global scale was established at the International Institute for Applied Systems Analysis (IIASA) through the pioneering work of Lutz et al. (2011; 2014) and was subsequently expanded and consolidated in collaboration with the European Commission Joint Research Centre (Lutz et al. 2018). The formulation of the population dimension of the Shared Socioeconomic Pathways (SSP; Kc and Lutz 2017), which embedded education as a core dimension of population heterogeneity across global development scenarios to 2100, cemented the study of the evolution of human capital as a central and indispensable concern of contemporary demographic research. This line of work was further updated in the most recent Wittgenstein Center (WIC) projection round, WIC2023, which revised the population and human capital components of the SSPs using a 2020 baseline and provides projections by age, sex, and level of education to 2100 (KC et al., 2024). The SSPs are a set of five global scenarios developed by the climate and integrated assessment modelling community that describe alternative trajectories of societal development, ranging from sustainability and international cooperation (SSP1) to fragmentation and fossil-fuel dependency (SSP5), and which embed contrasting assumptions about population growth, educational expansion, and urbanization (O'Neill et al. 2017). The present article aims to extend the existing human capital projections developed under the SSP scenarios to the sub-national level of European NUTS-2 regions. This will provide a more thorough understanding of how the twin transformations of educational expansion and demographic ageing can be expected to reshape Europe over the course of the 21st century.

The remainder of the paper is organised as follows. Section 2 provides the background to the study, first defining the conceptual and measurement boundaries of human capital, then reviewing the historical roots and contemporary dynamics of regional human capital disparities across Europe, and finally identifying the research gap this paper addresses. Section 3 presents the data sources, their harmonisation, and the projection framework. Section 4 reports the results across the SSP scenarios. Section 5 concludes.

2 Background

2.1 Defining and Measuring Human Capital: From Multidimensional Concepts to Education-Centred Approaches

The concept of human capital encompasses multiple dimensions, reflecting different disciplinary traditions and policy objectives. Human capital can be understood as the stock of capabilities, competencies, and knowledge that individuals and populations possess. These can be acquired through formal education, but also vocational training, and lived experience. Moreover, the definition of human capital also includes health, as the skills embodied by individuals are enhanced through physical and cognitive health (Lutz and Kc 2011). In the context of development and demographic analysis, several definitional approaches have emerged.

First, the multidimensional approach treats human capital as a composite of education, health, and social capabilities (OECD 2001). This way of thinking about human capital has led to the United Nations Development Programme's Human Development Index (HDI) for instance, integrating life expectancy at birth, mean years of schooling, expected years of schooling, and gross national income per capita to construct a holistic portrait of human development (UNDP, 2025). This approach acknowledges that health status, reflected in life expectancy and disability-adjusted life years, is an integral component of human functioning and economic participation. Similarly, the World Bank's Human Capital Index combines health and education measures to quantify the productivity of the next generation of workers (Kraay 2018).

Second, the education-centric approach isolates formal education as the primary metric of human capital stock and change. This tradition, rooted in Mincer (1958) and Becker (1993), and crystallised in contemporary demographic research through the work of Lutz and colleagues (Lutz and Goujon 2001; Kc et al. 2010; Lutz and Kc 2011; Goujon et al. 2016), treats educational attainment, measured by years of schooling or formal attainment levels, as the most directly observable, policy-relevant, and cross-nationally comparable proxy for the human capital endowment of a population. The rationale is multifaceted: formal educational attainment is measured consistently across countries and over time through international classifications (ISCED—International Standard Classification of Education); it is causally prior to labour market outcomes (Card 1999), health behaviours (Cutler and Lleras-Muney 2010), and civic engagement (Dee 2004); and it is directly manipulable through policy, making it the most actionable lever for human capital formation (Persson 2014; Brunello et al. 2016; Heckman et al. 2018).

For this analysis, I adopted the education-centric definition. This choice reflects both methodological pragmatism and theoretical considerations. Methodologically, education is a dimension of human capital for which long-run historical data and reliable national projections exist across Europe. Theoretically, education is often interpreted as a proxy for human capital, summarizing a bundle of capabilities—including cognitive skills, knowledge, and behavioural attributes—that underpin productivity and economic performance (Hanushek and Woessmann 2008; Psacharopoulos and Patrinos 2018). The focus on education allows us to trace the long-term regional evolution of human capital through a consistent metric, and to project forward the heterogeneous trajectories of educational attainment across NUTS 2 regions under alternative socioeconomic futures. It should be acknowledged that the focus on formal educational attainment disregards differences in actual skills at equal attainment levels. Recent work on skills-adjusted measures of human capital, such as the Skills in Literacy Adjusted Mean Years of Schooling (SLAMYS) presented in Lutz et al. (2021) has shown that accounting for skill differences can substantially alter cross-country comparisons of human capital stocks. However, projecting such skill adjustments credibly at the regional level remains beyond the reach of currently available data, and attainment remains the most robust foundation for long-range regional projections.

2.2 The Twin Transformation: Educational Expansion and Population Ageing in Europe

Human capital accumulation is shaped, in all advanced economies, by two powerful and partially countervailing forces: the expansion of educational attainment across successive cohorts and the structural ageing of populations (Lutz, Goujon, et al. 2008). The former raises the average level of skills and knowledge embodied by the workforce; the latter shifts the relative demographic weight of cohorts. Together, they determine the size, composition, and productive capacity of the human capital stock over time. Europe represents the most advanced laboratory for studying this dual dynamic simultaneously. The educational expansion has been one of the defining socioeconomic transformations of the post-war period in Europe. Tertiary attainment rates among young cohorts have risen from single-digit percentages in the 1950s to over 40 percent across the EU27 average today, with some regions (e.g., Île-de-France (FR), Sostinės regionas (LT)) approaching or exceeding 70 percent tertiary attainment among 25–34 year-olds (OECD 2023; Eurostat 2025a).

This expansion has been broadly inclusive, spanning both Western and Central-Eastern European regions, though at markedly different speeds and with distinct spatial patterns. Population ageing, by contrast, is a largely irreversible demographic phenomenon in Europe, driven by fertility rates that have fallen below replacement level (averaging 1.34 children per woman EU27-wide in 2024) combined with rising life expectancy (exceeding 80 years across most regions) (Eurostat 2025b, 2025c). According to Eurostat's latest population projections, the population considered to be of working-age at present (~20–64) is expected to decline in the majority of EU member states by 2050, while the share of individuals aged 65 and over is projected to increase from around 20% today to nearly 30% by mid-century; correspondingly, old-age dependency ratios are set to rise sharply (Eurostat 2024).

Yet the interplay between educational expansion and population aging is neither automatic nor uniform. While educational expansion generates upward pressure on average human capital levels, ageing mechanically reduces the size of the labour force from which skilled workers are drawn. At the regional scale, this interaction takes on particular urgency: the intensity of ageing varies dramatically across space (with Eastern and Southern European regions generally older, and faster-ageing, than Western and Northern regions), as does the pace of educational expansion (with urban areas generally expanding tertiary education more rapidly than rural areas (Rodríguez-Pose and Tselios 2011)).

2.3 The Long Shadow of History: Regional Persistence and Structural Inequality in European Human Capital

The regional heterogeneity evident in contemporary Europe does not emerge *de novo* but rather reflects deeply entrenched historical patterns of human capital formation that have structured European development for nearly two centuries. The empirical record reveals a striking degree of continuity in regional disparities. Hippe (2013), for example, shows that regions lagging behind in terms of human capital formation in the mid-nineteenth century remained largely the same ones that continued to lag in 1900 and even as late as 1930. Building on this long-run perspective, Hippe (2020), in his comprehensive historical analysis, traces the evolution of regional educational disparities from 1800 onwards, standardizing historical literacy, numeracy, and attainment data to modern NUTS regional boundaries. His findings reveal a persistent European core-periphery structure in human capital, with the most educated regions concentrated in the industrialized core of northwestern and central Europe and the peripheries of Ireland, Iberia, Southern Italy, and the Balkans lagging. Within national borders, too, deep historical divides are visible: the Italian North-South gap, the St. Malo–Geneva line in France, and the urban-rural cleavages in Eastern Europe all echo patterns already established in the 19th and early 20th centuries. Specifically, the correlation between regional literacy rates in 1900 and PISA mathematics performance scores in 2012 exceeds 0.77 (Hippe 2020), underscoring the path-dependent effects of historical human capital structures on contemporary educational attainment and acquired skills.

A key dimension of this persistence is that intranational inequalities in human capital have been as large as, and often larger than, international differences. Using numeracy, literacy, and educational attainment proxies for 102 European regions over 160 years, Diebolt and Hippe (2017) demonstrate that the within-country regional variance in human capital often exceeds the between-country variance, suggesting that a poorly educated region in a wealthy country may have substantially lower human capital than a well-educated region in a poorer country. This "subnational reality" has been largely obscured by nation-state-level analyses, yet it is essential for understanding regional development trajectories. Notably, this persistence extends beyond education alone: the same spatial patterns characterize long-run economic development across European regions (Hippe and Baten 2012), reinforcing the structural nature of these divides.

These patterns are not accidental or easily reversible. As argued in Diebolt and Hippe (2018, 2019), they reflect institutional arrangements, geographic constraints, and historical inequalities that took centuries to form and that have systematically limited convergence across regions. For contemporary policy, this means that observed regional disparities in human capital are deeply structurally embedded and unlikely to respond rapidly to short-term interventions.

2.4 Convergence and Divergence in Contemporary Regional Human Capital Dynamics

Shifting the focus from the long-run historical perspective to the more recent period, post mid-1990s evidence on regional human capital dynamics in Europe presents a nuanced picture: one of quantitative convergence in educational attainment levels combined with persistent and in some cases widening divergence in the capacity of regions to retain and benefit from their educated workforce. This duality has been explored in the literature from multiple disciplinary perspectives. In economic geography, Rodríguez-Pose and Tselios (2009, 2010, 2011) provide a series of foundational analyses of the spatial distribution and determinants of educational attainment and inequality across European regions. Drawing on microdata from the European Community Household Panel (ECHP) for 102 Western European regions, they identify persistent positive spatial autocorrelation, whereby highly educated regions tend to cluster together. Their findings also reveal a stable North–South and urban–rural divide in educational outcomes, a pattern corroborated by later studies (Chocholatá and Furková 2017). Studies of educational stratification further reveal that these spatial inequalities in attainment are not merely distributional but reflect and reinforce deeper patterns of social and economic disparity, persisting across generations in ways that standard convergence metrics do not fully capture (Betthäuser et al. 2021).

Dańska-Borsiak (2023) extends this line of inquiry by examining convergence dynamics in human capital across European NUTS 2 regions over the period 2005–2020. The study finds evidence of both absolute and conditional β-convergence, with Central and Eastern European (CEE) regions growing faster than their Western counterparts. However, the convergence process remains markedly slower in CEE regions, where the estimated half-life of human capital disparities is approximately eleven times longer than in Western Europe.

The picture becomes more complex, and more concerning, when the analysis extends beyond the creation of human capital to its deployment and retention, that is, the capacity of regional economies to absorb graduates into productive employment and to retain highly educated workers rather than losing them to more prosperous regions or countries. The Eurofound (2024) comprehensive report on the role of human capital inequalities in social cohesion and convergence finds that while EU Member States have converged upward on indicators of human capital creation (tertiary attainment, early school leaving, participation in lifelong learning), significant divergence persists in utilisation. Elevated rates of graduates who are neither in employment, education, nor training and systemic overqualification, predominantly concentrated in the peripheral regions of Greece, Spain, Italy, and parts of CEE, constitute evidence of human capital underutilisation that pure attainment indicators do not capture. A critical conceptual contribution of the Eurofound (2024) report is the identification of a 'talent-development trap': regions experiencing simultaneous demographic decline and educational stagnation enter a self-reinforcing downward spiral from which exit is structurally difficult.

The mobility dimension introduces a further layer of complexity. The Eurofound (2024) report documents that the economic loss from brain drain, the net emigration of highly educated graduates, is not compensated by returning or incoming talent in most net-losing Member States. Even countries that are net talent-receivers at the national level, such as Ireland and the Netherlands, contend with internal brain drain as human capital migrates from peripheral regions to primary urban agglomerations (Iammarino et al. 2019). Overall, the brain drain map can be said to be highly spatially uneven: in Spain it operates primarily nationally from inland regions to Madrid and the coast; in Lithuania, it operates internationally, typically to other EU member states (Pérez 2026). Left unaddressed, these subnational imbalances risk creating negative feedback loops that degrade local labour markets, weaken educational systems, and decelerate the EU's overall economic convergence.

These retention and mobility imbalances are further compounded by the structural ageing of regional populations. Prenzel and Iammarino (2021) examine how labour force ageing shapes the availability and composition of human capital across 324 German districts, finding that older workforces are systematically associated with lower aggregate educational attainment, as they tend to hold a higher share of vocational rather than tertiary qualifications. While rising attainment among younger cohorts has historically offset ageing effects at the national level, the authors show this balance to be increasingly fragile and spatially uneven. Regions with older demographic structures face particular difficulty attracting and retaining younger, more highly educated workers, thereby impeding the renewal of regional human capital stocks. Moving beyond a simple high/low-skill dichotomy, the effectiveness of human capital investment as a policy response to demographic change is fundamentally contingent on place-sensitive strategies that account for local industrial structures and demographic constraints.

2.5 The Research Gap and Scope of the Present Study

Projecting regional human capital requires moving along two dimensions simultaneously: from basic demographic variables to including educational attainment and from the national to the subnational scale. The literature has advanced along each dimension separately, but has yet to integrate them. At the national scale, the Wittgenstein Centre for Demography and Global Human Capital has produced the most comprehensive multi-scenario projections of population by age, sex, and educational attainment for over 200 countries to 2100 under the SSP framework (Lutz et al. 2014; EC. JRC 2018; KC et al. 2024). Loichinger (2015) extended this approach to labour force projections for 26 EU countries to 2053, while Marois et al. (2019, 2020) further advanced the tradition through microsimulation, modelling how changes in labour force participation, migration volume, and educational selectivity interact to shape future workforce quality and dependency ratios across EU member states. The QuantMig-Mic model (Potančoková et al. 2023) and, most recently, the Link4Skills framework (Marois et al. 2026) have pushed the frontier further by jointly projecting occupational distributions and skill-specific labour demand to 2060, revealing structural mismatches that aggregate national figures entirely conceal.

These national-level projections provide the benchmark for understanding Europe's human capital future, but they leave open the question of how these dynamics play out spatially. A parallel strand of literature has advanced the subnational dimension of demographic projection, though largely without the human capital component. The Eurostat produces deterministic, "what-if" demographic projections at the NUTS-3 level under a baseline scenario of partial convergence in fertility, mortality, and migration across EU and EFTA countries, complemented by five sensitivity variants: lower fertility, lower mortality, zero net migration, and higher and lower non-EU immigration (Eurostat 2024). While this represents the finest-grained, official, regional foresight instrument available for Europe, it does not disaggregate by educational attainment, leaving the human capital dimension of demographic change invisible.

These Eurostat projections have nonetheless provided the demographic foundation for Curtale et al. (2026), which proposed a regionalization methodology for downscaling GDP projections from the national to the NUTS 3 level, incorporating human capital — proxied by tertiary education shares — alongside demographic structure. Yet this approach treats educational attainment as an exogenous covariate, fixed in its relationship to economic output, rather than as an endogenously evolving dimension of the regional population. Critically, it is not linked to any scenario of educational expansion or differential skill development across regions; the spatial distribution of human capital is simply a static input, not a dynamic component of the projection itself.

Early subnational SSP-coherent population projections for European NUTS regions, produced by Terama et al. (2019), are limited to the age dimension and rely on constant-share allocation to distribute national totals downward, imposing static regional hierarchies that cannot capture spatially uneven demographic change. Again under the SSPs, Tamburini et al. (2025) developed a neural network-based architecture to disaggregate age-specific population projections to the NUTS-2 level for 34 European countries. Their results reveal substantial spatial heterogeneity in age structure at the subnational level, confirming that demographic change unfolds along deeply uneven within-country gradients. The grounding of this work in the SSP scenario framework, which explicitly embeds a human capital dimension, represents a valuable methodological step forward. However, the educational composition of the population was not yet carried through to the subnational disaggregation itself.

What remains largely absent from the literature, is a spatially explicit, forward-looking framework that projects human capital — understood as the joint product of educational attainment and demographic structure — at the subnational level in Europe. This gap matters for several reasons. First, as the historical literature by Hippe (2020) and Diebolt and Hippe (2019) makes clear, regional human capital trajectories are highly path-dependent and exhibit substantial intranational heterogeneity that national averages conceal. Second, as Dańska-Borsiak (2023), Prenzel and Iammarino (2021), and Eurofound (2024) show, the forces driving human capital accumulation — ageing, educational expansion, and migration — interact differently across regions in ways that aggregate projections cannot capture. Third, policy-relevant decisions in areas ranging from cohesion policy funding to long-term care infrastructure, education investment, and skill-development strategies require subnational foresight tools. The European Commission's own cohesion policy framework recognises regional heterogeneity as a central challenge (Iammarino et al. 2019), yet the demographic and human capital projections that inform it remain at the national level. This paper tries to fill this gap by developing a multi-dimensional, regional, human capital projection framework for European NUTS-2 regions, integrating the educational attainment and age-structure dynamics of regional populations under alternative demographic and policy scenarios. By doing so, it aims to produce the first systematic forward-looking account of how the dual transformation of educational expansion and demographic ageing will reshape the geography of human capital across Europe to mid-century, identifying the regions most at risk of human capital depletion, and those best positioned to sustain or augment their productive stock of skills and knowledge.

In this context, the present work introduces a Bayesian time series model designed to disaggregate SSP-consistent national projections into age-, sex-, and education-specific projections at the subnational level.

Building directly on the regional demographic projection framework of Tamburini et al. (2025), the addition of the educational attainment dimension represents a substantive methodological extension of that work: where Tamburini et al. (2025) established the subnational disaggregation architecture for age and sex structures, the present contribution augments this framework with a dedicated educational attainment module, enabling a fully integrated, uncertainty-aware treatment of human capital composition at the regional level. A central design principle of this extension is the explicit preservation of coherence with the Shared Socioeconomic Pathways (SSPs): subnational educational projections are constrained to remain consistent with the SSP national-level trajectories, ensuring that the regional disaggregation does not depart from the broader scenario logic that underpins comparable work in climate, energy and integrated assessment modelling. This coherence is essential for the downstream applicability of the results, particularly in the context of linking human capital dynamics to assessments of climate vulnerability and labour force participation. My goal is therefore to deliver projections that are simultaneously internally consistent at the subnational scale, anchored to globally recognised scenario frameworks, and granular enough to support the spatial analysis necessary to understand and manage the evolution of human capital across European regions until 2100.

3 Data and Model

3.1 The SSPs and Their Education Dimension

The main input to the analysis conducted in this research comes from the Shared Socioeconomic Pathways (SSPs). The SSPs constitute the standard global scenario framework for integrated research on climate change, demographic change, and socioeconomic development (O'Neill et al. 2014; Riahi et al. 2017). Originally developed to span a wide range of plausible futures along two dimensions — challenges to mitigation and challenges to adaptation — the five SSPs combine qualitative narratives with quantitative projections of population, economic growth, and urbanisation to 2100 (O'Neill et al. 2017). Their demographic and human capital dimension was formalised by Kc and Lutz (2017), who applied multi-state cohort-component methods to project national populations by age, sex, and educational attainment under each SSP storyline — a contribution subsequently updated and extended to 201 countries by Lutz et al. 2018 and most recently revised KC et al. 2024. While originally designed as broad socioeconomic storylines, each pathway implies distinct trajectories of educational expansion, human capital formation, and inequality in access to education that are key determinants of adaptive capacity, labour productivity, and long-run economic development (Lutz et al. 2008; Lutz et al. 2014).

SSP1 (Sustainability) assumes rapid and equitable educational expansion globally, underpinned by strong institutional capacity and inclusive policies that substantially reduce barriers to schooling; transition rates to higher education follow an average of the historical fast-track and global education trend trajectories, such that by 2050 the global mean years of schooling would reach approximately the current European level (Kc and Lutz 2017).

SSP2 (Middle of the Road) projects a continuation of recent historical trends, with moderate and heterogeneous improvements in enrolment and graduation rates across regions, reflecting persistent but slowly narrowing gaps in access and attainment; it represents the median trajectory and serves as the reference scenario in many applied studies (Lutz et al. 2018; Rao et al. 2019).

SSP3 (Regional Rivalry) describes a world of fragmentation and stalled development in which weak international cooperation constrains investment in human capital: educational expansion halts or reverses in many regions, fertility remains high, and strong disparities in attainment persist both across and within countries, producing the largest global population by 2100 and the lowest aggregate educational attainment of all five pathways (Kc and Lutz 2017; Fujimori et al. 2017).

SSP4 (Inequality) is distinctive in that it does not assume uniformly low or high education but rather a deeply polarised distribution: a well-educated, internationally connected elite coexists with large shares of the population holding only primary or lower secondary qualifications, generating high within-country educational inequality alongside moderate aggregate attainment levels and reinforced intergenerational stratification (Calvin et al. 2017; Kc and Lutz 2017)

SSP5 (Fossil-fuelled Development), like SSP1, assumes rapid educational expansion and substantial human capital investment globally, driven by strong economic growth and technological change; the two scenarios are therefore similar in their educational trajectories but diverge sharply in their energy and emissions pathways (Riahi et al. 2017; O'Neill et al. 2017).

Across all five pathways, educational attainment is not merely an output of the scenario but an active driver of projected demographic behaviour: fertility, mortality, and migration differentials by education level are built into the multi-state projection model, such that the educational scenarios generate internally consistent demographic trajectories rather than simply overlaying attainment assumptions onto a fixed population structure (Lutz and Goujon 2001; Kc and Lutz 2017). These differences in educational trajectories critically shape long-term demographic structure, labour productivity, and adaptive capacity, making education a central variable in any integrated application of the SSP framework and, by extension, a prerequisite for any scenario-based projection of human capital at the subnational level.

In the European context specifically, the inter-scenario variation in human capital trajectories is narrower than at the global level, reflecting the already high baseline of educational attainment among the continent's population and, consequently, the limited scope for further dramatic expansion in the most optimistic pathways. The more consequential differentiation for Europe operates along two other axes: (1) the pace at which remaining gaps between Central and Eastern European regions and their Western counterparts close under different scenarios, and (2) the degree to which population ageing compresses or erodes the human capital gains generated by continued educational expansion, a dynamic that varies substantially across SSPs and whose subnational dimension remains, to date, largely unexplored.

3.2 The Empirical Data

While the SSPs provide information under different, scenario-specific assumptions, the empirical part of the analysis draws on three complementary data sources:

1. Eurostat Database[1]: provides NUTS-2 regional population estimates (in thousands) disaggregated by sex, five-year age group, and educational attainment level, with attainment reported according to grouped ISCED levels: low (ISCED 0–2), medium (ISCED 3–4), and high (ISCED 5–8). I assembled a consistent panel for 21[2] countries covering all corresponding NUTS-2 regions (Eurostat 2025a), except for Åland (FI20) and France's overseas departments and Corsica due to missing data (for a detailed overview of coverage, see Table 1 in the Appendix). Despite some gaps and sampling variability, particularly in smaller regions and earlier years, this remains the most harmonised and comprehensive source of regional educational attainment available at the European level, and it serves as the historical input for the modelling framework.

[1] https://ec.europa.eu/eurostat/databrowser/view/lfst_r_lfsd2pop/default/table?lang=en
[2] Austria (AT), Belgium (BE), Bulgaria (BG), Cyprus (CY), Czechia (CZ), Germany (DE), Denmark (DK), Estonia (EE), Spain (ES), Finland (FI), France (FR), Greece (GR), Croatia (HR), Hungary (HU), Ireland (IE), Italy (IT), Lithuania (LT), Luxembourg (LU), Latvia (LV), Malta (MT), Netherlands (NL), Poland (PL), Portugal (PT), Romania (RO), Sweden (SE), Slovenia (SI), Slovakia (SK)

2. Wittgenstein Centre Human Capital Data Explorer (WCDE)[3]: this source supplies SSP-coherent national population trajectories by age, sex, and educational attainment for the period 2020–2100, covering all five Shared Socioeconomic Pathways (KC et al. 2024). The WCDE provides the national-level demographic and educational scenario assumptions onto which the subnational projection is anchored. The WCDE reports educational attainment in six categories (No Education, Incomplete Primary, Primary, Lower Secondary, Upper Secondary, and Post Secondary), which were aggregated into the three ISCED-based groups used throughout this study to ensure consistency with the Eurostat regional data: No Education, Incomplete Primary, Primary, and Lower Secondary were combined into the low-education group (ISCED 0–2); Upper Secondary was mapped to the medium group (ISCED 3–4); and Post Secondary to the high group (ISCED 5–8). This harmonization step was necessary because the WCDE and Eurostat classifications do not align directly, and a common three-category scheme was required to anchor the subnational projections to the national WCDE trajectories.

3. Tamburini et al. (2026): NUTS-2 level SSP-coherent population projections by age and sex: this dataset provides age and sex specific population projections for 305 NUTS 2 regions across 34 European countries over the 2020-2100 period. Projections align with the latest SSP country scenarios and include breakdowns by sex and five-year age groups. Regional age structures are projected using a neural network architecture, followed by an SSP consistent sex split that incorporates country level trends while retaining regional heterogeneity.

3.3 Modelling approach

Time series models have long played a central role in demography, providing a framework to analyse and forecast the temporal evolution of key indicators such as mortality, fertility, and population size. Classical approaches typically rely on stochastic processes, as autoregressive integrated moving average (ARIMA) models or latent factor models like Lee–Carter, to capture trends, seasonality, and temporal dependence in demographic rates. While effective, these approaches are often estimated in a frequentist setting and may treat parameters as fixed, limiting their ability to fully propagate uncertainty, especially when data are sparse, noisy, or heterogeneous across populations. In response to these limitations, Bayesian time series methods have gained increasing prominence in demographic research. By treating model parameters as random variables and combining prior information with observed data, Bayesian approaches enable coherent uncertainty quantification and probabilistic forecasting (Bijak and Bryant 2016; Bryant and Zhang 2018).

[3] https://dataexplorer.wittgensteincentre.org/wcde-v3/

Moreover, hierarchical Bayesian time series models allow for the pooling of information across countries, sexes, or regions, improving estimation in data-limited contexts while preserving heterogeneity (Alkema et al. 2011; Wiśniowski et al. 2015). These models are particularly well-suited for demographic applications, where multiple sources of uncertainty as sampling error, measurement error, and structural variability must be jointly addressed. This paradigm underpins the now-standard probabilistic population projections produced by the United Nations, which rely on Bayesian hierarchical time series models for key demographic components such as fertility and mortality (Raftery et al. 2014). As a result, Bayesian time series frameworks have become a cornerstone of modern population forecasting, underpinning recent advances in global mortality, fertility, and population projection models (Keilman 2020; Chao et al. 2021).

Forecasting education-related variables poses distinct methodological challenges relative to other demographic or economic projections. Unlike mortality or fertility, which respond to biological constraints and are routinely embedded in the cohort-component projection framework, educational attainment is shaped by a combination of individual life-course decisions, institutional structures, policy interventions, and intergenerational transmission processes that are inherently difficult to model in a parsimonious and generalisable way (Barro and Lee 2013; Lutz and Goujon 2001). The literature has consequently developed along two relatively separate tracks: a macro-level tradition focused on projecting aggregate educational attainment distributions across populations, and a micro-level tradition concerned with modelling individual educational transitions and their determinants.

Within the micro-level tradition, Mao et al. (2024) provide a systematic and comprehensive exploration of time series analysis as a pivotal tool for data-driven decision-making and personalised learning within the field of education research. Addressing previously fragmented research, the authors establish a detailed taxonomy of educational data, categorising it into academic performance, learning behavioural, socio-demographic, and psychosocial types collected from in-person, online, and blended environments. The review synthesises the application of four core methodologies: forecasting, classification, clustering, and anomaly detection. This demonstrates their utility in critical educational tasks such as predicting student success, profiling learner behaviours, and identifying at-risk students. Beyond traditional methods, the review highlights emerging directions, including the transformative potential of multimodal data fusion and the integration of large language models (LLMs) for more adaptive and equitable learning environments. Other works have focused on the prediction of enrolments, teacher statistics, and school expenditures (S. Yang et al. 2020; Sinuany-Stern 2021; James and Weese 2022) or on the evaluation and improvement of common forecasting methods for educational planning (Tang and Yin 2012; Tang and Chou 2016).

Within the macro-level tradition, the dominant approach has been the multi-state cohort-component model, in which populations are stratified by educational attainment level alongside age and sex, and projected forward through age- and education-specific rates of transition (between education levels), fertility, mortality, and migration. Applications at the subnational level have, however, remained limited, either to specific areas and scenarios (Emelyanova 2019) or to single countries characterised by high degrees of internal heterogeneity (Goujon and McNay 2003; Goujon et al. 2000; S. K.C. et al. 2017; Kc et al. 2018). A long-standing example of this approach at the national level is the series of education projections produced by the US National Center for Education Statistics (NCES) since 1964. The most recent edition, Projections of Education Statistics to 2030 (Irwin et al. 2024), generates projections using exponential smoothing for enrolments and graduates, and linear regression for teaching staff and expenditures, drawing on historical data extending in some cases to 1869–70. The framework operates at national, regional, and state level, covering all 50 states and the District of Columbia, with a projection horizon of ten years and no alternative scenarios.

Against this backdrop, the present work departs from both the cohort-component tradition and the standard frequentist time series approaches reviewed above, proposing instead a Bayesian time series framework specifically designed to handle the compositional and multi-dimensional nature of educational attainment distributions at the subnational level. Rather than projecting educational attainment from scratch, the framework is designed to disaggregate existing national-level educational projections down to the subnational level, leveraging projected age- and sex-specific population structures as an anchoring scaffold. The methodological choices underlying this framework are motivated by the specific structure of the data and the projection objectives.

Achieving this requires a flexible statistical framework capable of capturing relative changes in category proportions while respecting the compositional nature of the data and maintaining consistency with the national-level trajectories into which the subnational estimates must aggregate. To this end, rather than working with level-based latent processes as in traditional approaches, I modelled log-returns, defined as first differences of log-transformed shares, which offer a natural way to capture such relative changes. Log-transformations are widely used to stabilize variance and linearize multiplicative dynamics in time series (Box and Cox 1964; Tsay 2010), while log-differences often mitigate heteroskedasticity in proportion data. Moreover, log-returns are additive over time, facilitating the use of autoregressive structures in a transformed space (Tsay 2010). This approach is closely related to log-ratio methods from compositional data analysis, which map simplex-constrained data to an unconstrained space while preserving relative information (Aitchison 1986; Egozcue et al. 2003). Back-transformation via exponentiation and normalization ensures that predicted values remain valid compositions (Kynčlová et al. 2015). These properties make log-return representations particularly well-suited for dynamic compositional modelling of educational attainment distributions.

Building on this data transformation, I employed a Bayesian time series modelling framework. This class of models was chosen because of its ability to explicitly quantify uncertainty in the input data, which in this case, are derived from survey sources and therefore subject to sampling variability and measurement error. Beyond uncertainty quantification, Bayesian time series models also allow the incorporation of prior information, hierarchical structures, and flexible parameter estimation, making them particularly suitable for cross-country and multi-dimensional applications. The chosen specification extends a latent-trend multinomial framework into the log-return space, capitalizing on the benefits of log-returns for capturing relative changes while introducing stratification by country (N), sex (S), and scenario (P). After the usual MCMC posterior sampling of predictive counts from the multinomial model, I applied a correction procedure to align the results with two key sets of external marginal constraints—regional population projections (by country, region, sex, age, and scenario) and national education category projections (by country, sex, category, age, and scenario). This adjustment was implemented using a two-dimensional Iterative Proportional Fitting (IPF) algorithm enhanced with entropy regularization, ensuring internal consistency while preserving the statistical structure of the sampled distributions. The resulting structure is robust to heterogeneous regional dynamics and varying starting years across countries, ensuring greater comparability and reliability of trend estimates across diverse demographic and geographic contexts.

Notation

Symbol	Description
N	Number of countries
R_n	Number of regions in country n
S	Number of sexes
T	Number of time points
A	Number of age groups
C	Number of education categories (fixed at 3)
P	Number of projection scenarios
$t_{init}[n]$	Initial observation year for country n
$log\ p_{n,r,s,c,t,a,p}$	Log-probability for region within country, sex, edu.att, year, age, scenario
$logreturn_{p,n,r,s,c,t,a,p}$	First-order difference of log-probabilities[4]
$N_{n,r,s,t,a,p}$	Region's population by age, sex, year, scenario
$N^{1000}_{n,r,s,t,a}$	Regional population by age, sex, year and scenario (multinomial denominator fixed at 1,000)
$y_{n,r,s,t,a}$	Observed pop. size by region, age, sex, year, edu.att.
$y^{1000}_{n,r,s,1:3,t,a}$	Observed pop. Size values expressed per mille (‰); the sum across edu. att. categories within each region, age, sex and year equals 1000.
$PoPnat_{n,s,c,t,a,p}$	National level population by age, sex, year and scenario and edu. att.

Time Rescaling

Each country's time series is rescaled to [0, 1] to support cross-national trend pooling:

4

$$logreturn_{p,n,r,s,c,t,a} = \frac{log\ p_{p,n,r,s,c,t,a}}{log\ p_{p,n,r,s,c,t-1,a}} = log\ p_{p,n,r,s,c,t,a} - log\ p_{p,n,r,s,c,t-1,a}$$

$$time_{scaled_{n,t}} = \frac{t - t_{init}[n]}{T - t_{init}[n]} \tag{1}$$

Priors

I placed country- and sex-specific priors on both latent initial states and return dynamics:

$$\tau\_init_{n,s,a} \sim \Gamma(1,0.1) \tag{2}$$

$$\tau\_return_{n,s,a} \sim \Gamma(1,0.1) \tag{3}$$

$$\mu\_\beta_{n,s,c,a} \sim \mathcal{N}(0,1) \tag{4}$$

$$\tau\_\beta_{n,s,c,a} \sim \Gamma(1,0.1) \tag{5}$$

$$\beta_{n,r,s,c,a} \sim \mathcal{N}\left(\mu\_\beta_{n,s,c,a}\ ,\ \tau\_\beta^{-1}{}_{n,s,c,a}\right) \tag{6}$$

Latent Process in Log-Return Space

Initialization

At the initial year $t = t_{init}[n]$, log-probabilities are initialized as:

$$\log p_{n,r,s,c,t_{\text{init}}[n],a} \sim \mathcal{N}\left(0, \tau\_init_{n,s,a}^{-1}\right) \tag{7}$$

Dynamic Evolution

From $t = t_{init}[n] + 1$, onward, log-returns are drawn and accumulated additively:

$$logreturn_{n,r,s,c,t,a} \sim \mathcal{N}\left(\beta_{n,r,s,c,a}\ , \tau_{\text{return},n,s,a}^{-1}\right) \tag{8}$$

$$\log p_{n,r,s,c,t,a} = \log p_{n,r,s,c,t-1,a} + \text{log-return}_{n,r,s,c,t,a} \tag{9}$$

Probability Transformation and Likelihood

After computing log-probabilities for c = 1, 2, the third category is derived via normalization:

$$\alpha_{n,r,s,1,t,a} = 1 \tag{10}$$

$$\alpha_{n,r,s,2,t,a} = \exp\left(\log p_{n,r,s,1,t,a}\right) \tag{11}$$

$$\alpha_{n,r,s,3,t,a} = \exp\left(\log p_{n,r,s,2,t,a}\right) \tag{12}$$

$$Z_{n,r,s,t,a} = \sum_{c=1}^{3} \alpha_{n,r,s,c,t,a} \tag{13}$$

$$p_{n,r,s,c,t,a} = \frac{\alpha_{n,r,s,c,t,a}}{Z_{n,r,s,t,a}} \quad (14)$$

$$y^{1000}_{n,r,s,1:3,t,a} \sim Multinomial\left(p_{n,r,s,1:3,t,a}, N^{1000}_{n,r,s,t,a}\right)$$

where $N^{1000}_{n,r,s,t,a}$ denotes the population total rescaled to a fixed denominator of 1000 across all regions, ensuring that the multinomial likelihood captures the compositional structure of educational attainment independently of regional population size.

In forward projections where the total population varies by scenario:

$$pp_{n,r,s,c,t,a,p} = p_{n,r,s,c,t,a} \cdot N_{n,r,s,t,a,p}$$

This formulation preserves model uncertainty while enabling scenario-based benchmarking. The sex and country dimensions enable a specific trend estimation. The different future scenarios, namely SSPs 1 to 5, are directly involved in the model estimation from the first projection step.

Posterior draws IPF

To ensure coherence with the regional (Tamburini et al. 2026) and national level (KC et al. 2024) projections of the SSPs while maintaining the uncertainty quantification typical of the Bayesian model fitting, the posterior sampling results are processed through a correction procedure for the predicted counts from the multinomial model $pp_{n,r,s,c,t,a,p}$. This is a two-dimensional Iterative Proportional fitting (Fienberg 1970) algorithm.
Since the marginal totals $N_{n,r,s,t,a,p}$ and $PoPnat_{n,s,c,t,a,p}$ are SSP scenario (*p*) coherent, the IPF procedure inherently enforces scenario alignment for the posterior estimates. Each posterior sample is thus reweighted per-scenario, ensuring internal consistency of subnational projections with national-level assumptions across all dimensions: country, sex, age, and scenario.

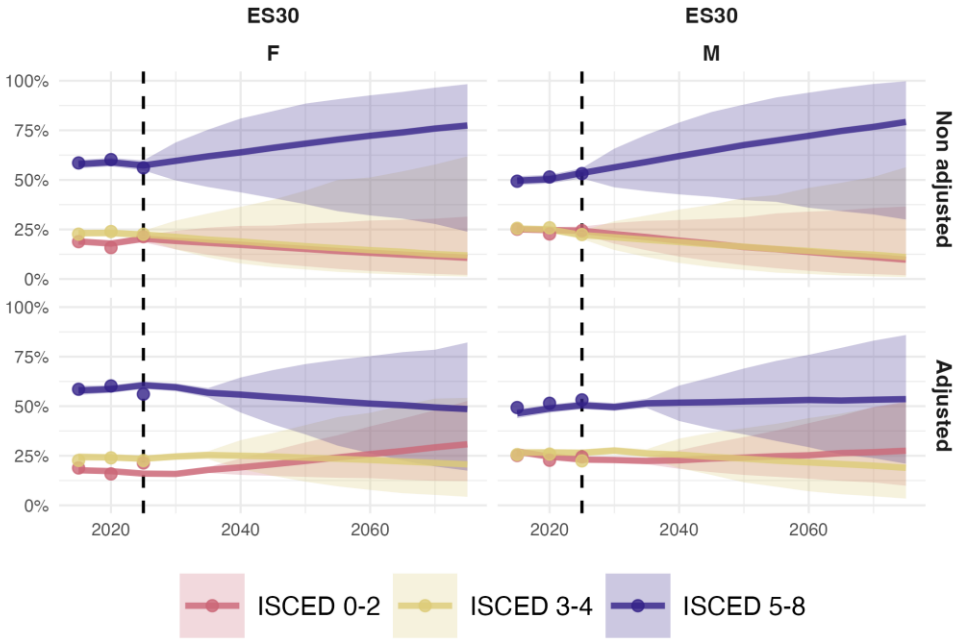

Figure 1: Effect of the IPF correction on posterior predictive distributions for the Madrid region (ES30), age group 25-44, under SSP-3. Each panel shows the median projection (line) and 95% credible interval (shaded area) for the three educational attainment categories (ISCED 0-2, ISCED 3-4, ISCED 5-8), separately by sex (F, M) and correction status (non-adjusted, adjusted). Dots represent observed input values. The dashed vertical line marks 2025, the end of the observed period.

Figure 1 illustrates the effect of the IPF correction on the posterior predictive distributions for a representative case. The most immediate consequence of the adjustment is a narrowing of the credible intervals, which is particularly pronounced in the near-term projection horizon. This is expected: in the years immediately following the observed period, the projected trajectories remain close to the empirical trends, and the IPF constraints, which enforce consistency with national and regional marginal totals, act as additional information that reduces posterior uncertainty. As the projection horizon extends and the SSP scenario assumptions begin to exert a stronger influence on the trajectories, the projected shares may deviate more substantially from the patterns observed in the historical data. In these later periods, the scenario-driven divergence introduces genuine compositional uncertainty that the IPF procedure cannot fully absorb, and the credible intervals consequently widen again.

3.3 Model fit and performance

Posterior distributions of model parameters and educational attainment shares were sampled via Markov Chain Monte Carlo within JAGS (Plummer 2003), implemented through the `rjags` package in `R` (R Core Team 2021), following a burn-in period of 5,000 iterations and a total of 50,000 iterations. Convergence was assessed using the Gelman-Rubin diagnostic (Gelman and Rubin 1992) and visual inspection of trace plots across chains, with all key parameters showing satisfactory convergence, $\hat{R}$ values below 1.1 and stable, well-mixed traces. The model was estimated on a MacBook Pro equipped with an Apple M1 Pro chip, 8 cores (6 performance and 2 efficiency), and 16 GB of unified memory; as a reference for computational cost, the model for Belgium, comprising 11 NUTS-2 regions close to the sample mean, required approximately 29.8 minutes to complete.

As a first validation step, I assessed the calibration of the model's posterior predictive uncertainty by means of a held-out forecasting exercise. The model is trained on observations up to and including 2020, and posterior predictive distributions are then generated for the five-year horizon spanning 2021 to 2025. For each region and educational category, I recorded whether the observed value falls within the posterior credible interval at four nominal coverage levels and compared the empirical coverage to the nominal target. Table 1 reports the results. The model performs well at lower confidence levels, with empirical coverage closely tracking nominal coverage, and shows a modest but consistent undercoverage at higher confidence levels. Overall, the results indicate that the model produces well-calibrated uncertainty intervals across the projection horizon, providing a reliable basis for the subsequent regional projections.

% CI	**% contained in the CI**
85	83
90	87
95	91
99	95

Table 1: Posterior predictive coverage of the held-out period 2021-2025. For each nominal credible interval level, the table reports the empirical percentage of observed regional educational attainment shares falling within the corresponding posterior credible interval, estimated from a model trained exclusively on data up to and including 2020. Note: coverage is computed across all NUTS-2 regions and educational attainment categories included in the projection framework.

As a second validation step, I verified the internal coherence of the results by comparing aggregated projections to reference distributions. Since no direct external benchmark exists at the NUTS-2 level for educational attainment by age and sex simultaneously, validation necessarily proceeds through reaggregation: I compared projections aggregated to the national level by educational attainment, as well as projections reaggregated at the regional level by age and sex, against the corresponding reference distributions, computing absolute percentage errors (APE) in both cases.

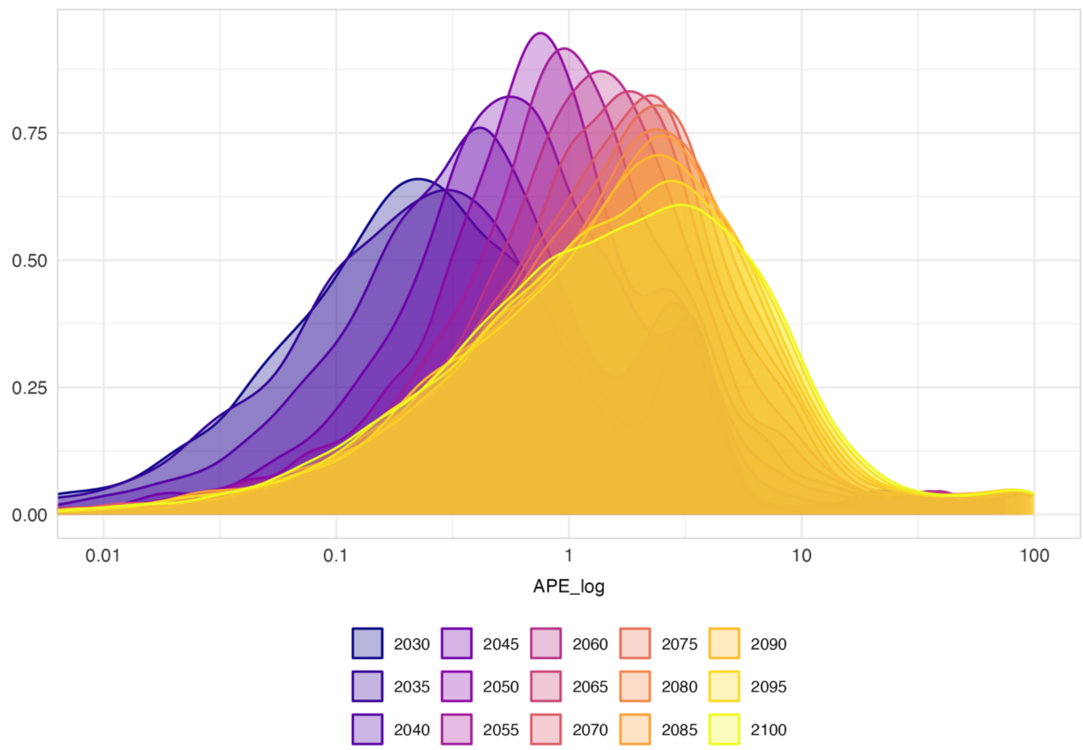

Figure 2: Distribution of the Absolute Percentage Error (APE, %) between NUTS-2 level projections and regional reference values, computed across all five SSP scenarios, three age groups (25–44, 45–64, 65+) and both sexes. The x-axis is displayed on a $\log_{10}$ scale, with tick labels showing the original APE values.

Figure 2 displays the distribution of the Absolute Percentage Error (APE) between the reaggregated NUTS-2 projections and the reference values, across all SSP scenarios, age groups and sex combinations, for the period 2030–2100. The x-axis is on a log_{10} scale to accommodate the highly right-skewed nature of the error distribution.

For the nearest horizon (2030, dark purple), the distribution is concentrated around values below 1%, indicating that the reaggregation procedure performs well in the short term. However, a clear and systematic rightward shift of the distribution is observed as the projection horizon increases, with the 2100 distribution (yellow) exhibiting a long right tail extending beyond 10% APE for a non-negligible share of region-scenario-demographic combinations.

This pattern is expected and reflects the compounding of uncertainty over time: errors in the underlying national-level projections and in the IPF correction procedure accumulate across projection periods, leading to progressively larger discrepancies at the NUTS-2 level. Notably, the bulk of the distribution remains below 2% APE even for the longest horizons, suggesting that the reaggregation procedure is broadly reliable, while the long right tail points to a subset of regions or demographic groups where the correction performs less well, like in small regions with volatile age-education structures.

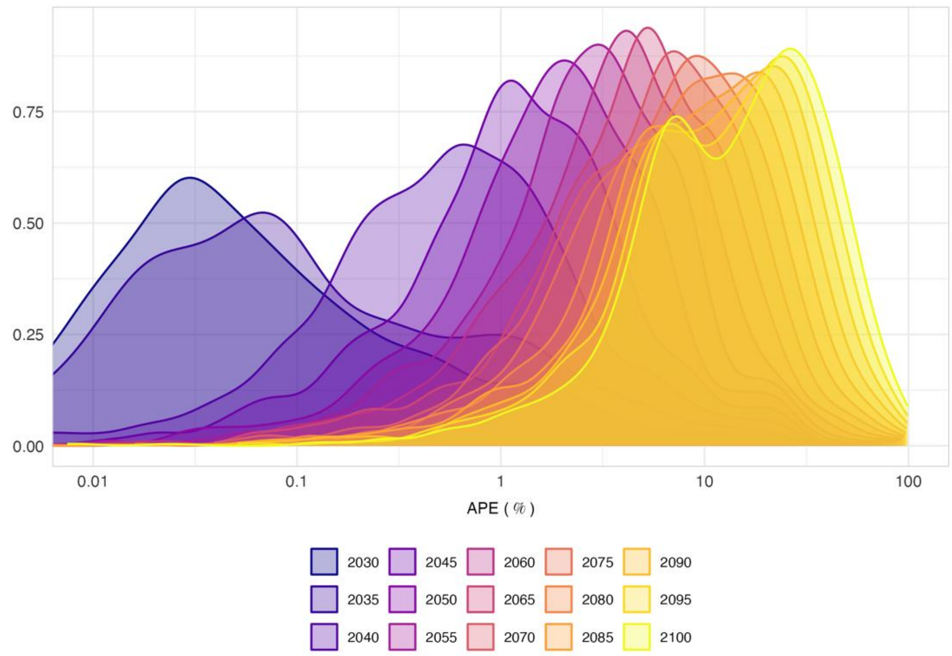

Figure 3: Distribution of the Absolute Percentage Error (APE, %) between reaggregated NUTS-2 projections and national-level reference values from the WCDE. The x-axis is displayed on a $\log_{10}$ scale, with tick labels showing the original APE values. Results are shown across all five SSP scenarios, three age groups (25–44, 45–64, 65+), three education levels (ISCED 0–2, ISCED 3–4, ISCED 5–8) and both sexes.

Figure 3 presents the distribution of the Absolute Percentage Error (APE) between the reaggregated national projections and the WCDE reference values, across all SSP scenarios, age groups, sex and educational attainment categories for the period 2030–2100. At short horizons (until 2050), the distribution is concentrated below 0.2% APE, indicating that the reaggregation procedure closely tracks the reference values when projections remain anchored to historically observed trends. As the horizon extends, the distributions shift progressively rightward, with the 2100 distribution centred around 5 to 10% APE. This deterioration reflects the compounding divergence between the SSP-specific assumptions embedded in the WCDE national projections and the regionally extrapolated trends from the Bayesian model: as scenarios such as SSP3 impose increasingly sharp departures from recent empirical trends in education expansion and demographic dynamics, the reaggregated estimates naturally diverge from the national targets. Nevertheless, even at the longest horizons the bulk of the distribution remains below 10% APE, suggesting that the procedure preserves national level constraints reasonably well throughout the projection period. This consistency across hierarchical levels and dimensions supports the plausibility of the model outputs and increases confidence in their use for scenario-based population analyses.

4 Results

From the resulting median of the posterior distribution, and once the projection had been completed over the full horizon to 2100, the absolute values were adjusted through a final round of IPF to fit the marginals given by the country-specific age group, sex, and educational attainment population projections, and the NUTS-2 age group- and sex- specific projections described in Section 3.2. To prevent implausible sex compositions at the regional level, I constrained the female-to-male ratio within each NUTS-2 region, age group, and educational attainment cell to the interval [0.4, 3.3]. These bounds were obtained by respectively augmenting and reducing by 25% the 2.5% and 97.5% quantiles of the observed age group and educational attainment sex ratios for the NUTS-2 regions. These substantially exceed the range of sex ratios observed in European adult populations, which for specific education categories rarely departs from near parity by more than a factor of two (Eurostat, 2023). Values outside this range in the historical data are attributable to small cell sizes in sparsely populated NUTS-2 regions rather than genuine demographic phenomena and are not expected to constitute plausible projection outcomes.

The resulting dataset constitutes a direct extension of Tamburini et al. 2026, augmenting the original NUTS-2 population projections by age group and sex with an educational attainment dimension. It covers 21 European countries across three key adult age groups: 25 to 44, 45 to 64, and 65 and older, under all five SSP scenarios. For a detailed overview of data coverage, see Table A1 in the Appendix. The projection results indicate a consistent and spatially pervasive improvement in educational attainment across European NUTS-2 regions: as illustrated in Figure 4, the share of the low-education group (ISCED 0–2) declines across virtually all regions between 2025 and 2075 under SSP2, with particularly pronounced reductions in Southern and Eastern Europe where low attainment was historically most prevalent, while the share of the tertiary-educated population (ISCED 5–8) expands across the continent, albeit with considerable variation in magnitude reflecting the heterogeneous starting conditions and structural characteristics of regional labour markets and education systems. These general improvements present different dimensions of variation. Within the same country, education improvement follows different trajectories across regions.

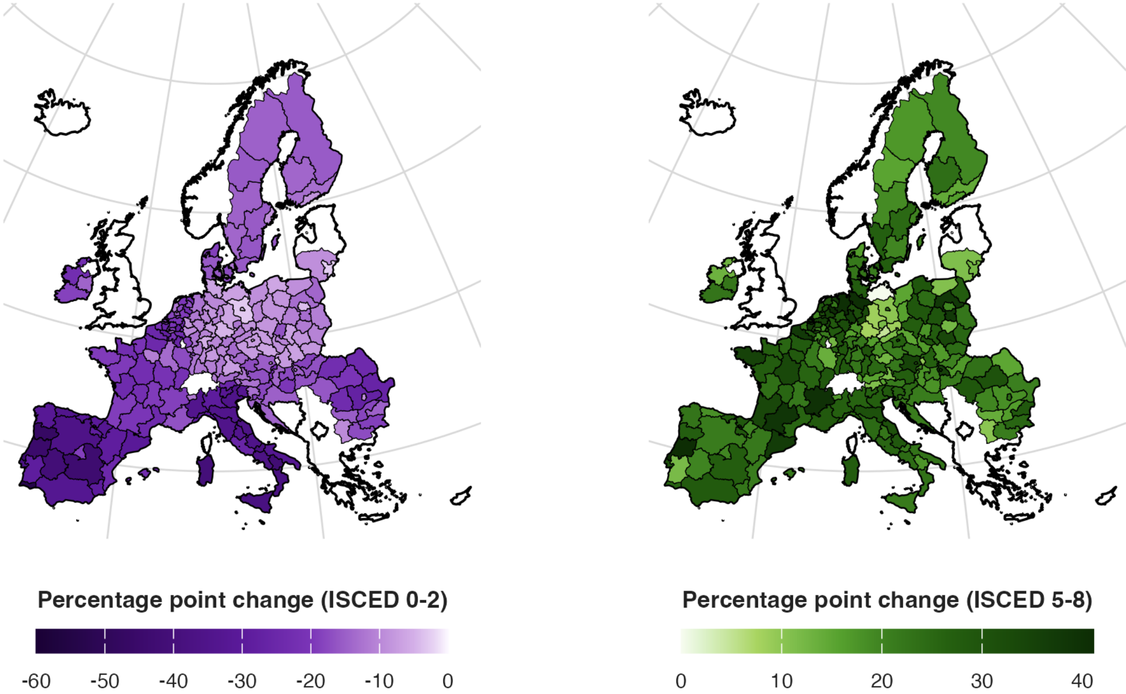

Figure 4: Percentage point change between 2025 and 2075 in the share of the population aged 25 and above with low educational attainment (ISCED 0–2, left panel) and tertiary education (ISCED 5–8, right panel) across European NUTS-2 regions under SSP2.

Figure 5 presents the evolution of the share of the population with tertiary education (ISCED 5-8) under SSP 2 for all NUTS-2 regions in each of the 21 countries analysed. The orange line represents the NUTS-2 region where the capital city is located, highlighted as it typically concentrates the highest levels of educational attainment among the working age population. A general upward trend is visible across all countries, with the capital region's share consistently and often remarkably higher than that of other regions. This is particularly pronounced in countries dominated by a single large metropolitan area, such as Austria, Hungary, Denmark and Czechia, where the gap between the capital and the rest of the country is both large and persistent over the projection horizon. An interesting divergence emerges for several former communist countries, like Slovakia, Romania, Bulgaria and Poland, where the share of highly educated individuals was already comparatively high at the start of the observation period, resulting in a flatter trajectory as other regions gradually converge toward similar levels. Some specific slightly decreasing or stagnating trends can be identified for the Azores region in Portugal and for the former East German regions, reflecting the historically observed demographic and educational dynamics in these areas.

Resulting from the specific modelling approach, the projected trajectories are strongly shaped by the educational attainment trends observed at the regional level during the observation period 2005 to 2025, with future directions mediated by the national level human capital evolution implied by each SSP scenario. Nevertheless, the general education expansion tends to compress the relative advantage of capital city regions and other historically high-education areas. As tertiary attainment rises across all regions, the share of highly skilled individuals concentrated in the leading regions gradually declines relative to the national distribution, suggesting a process of educational convergence within countries across the working age population (age groups 25 to 44 and 45 to 64). This dynamic is documented in greater detail in Figure A1 in the Appendix.

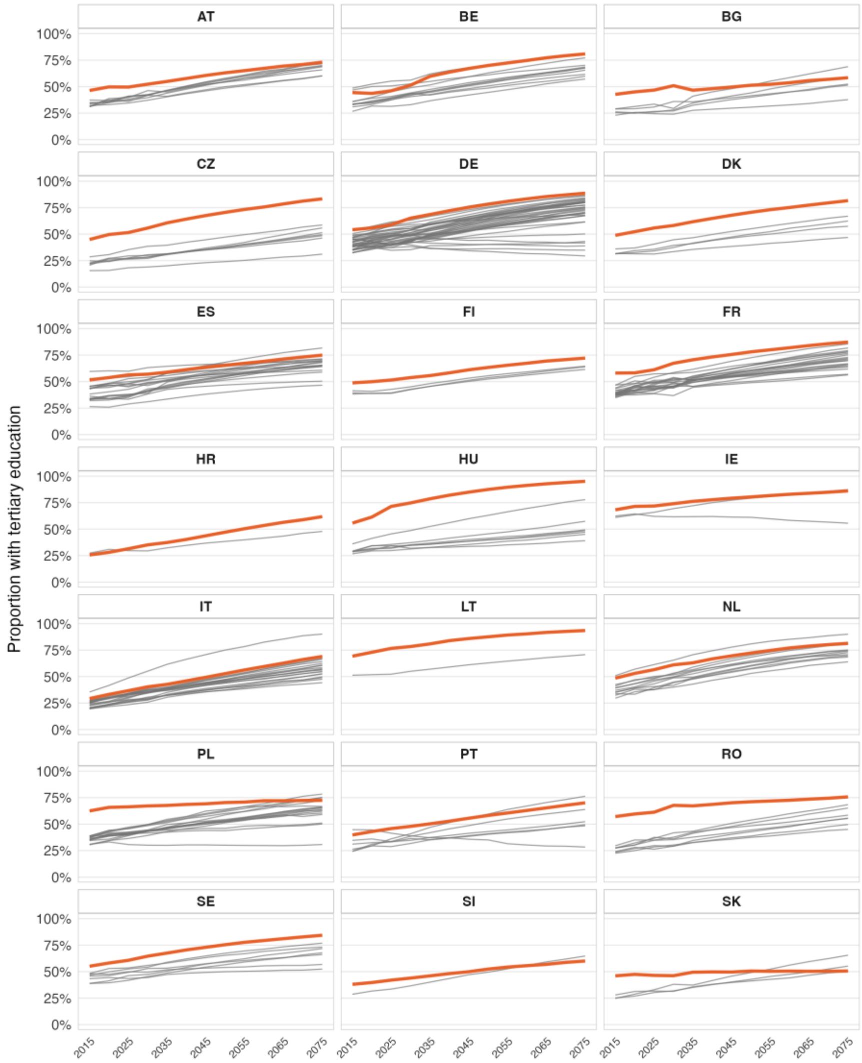

Figure 5: Projected share of the population aged 25 to 44 with tertiary education (ISCED 5-8) at the NUTS-2 level under SSP scenario 2, for both sexes combined, from 2015 to 2075. Each grey line represents a single NUTS-2 region within the country; the orange line highlights the capital region.

The aggregate educational expansion documented above, however, conceals important heterogeneity along the sex dimension: while both men and women participate in the projected growth of tertiary attainment, the pace and magnitude of these gains differ systematically, with women consistently outpacing men across regions, age groups and projection horizons. Figure 6 presents the distribution of the female-to-male (F/M) ratio in tertiary educational attainment (ISCED 5–8) across NUTS-2 regions under SSP3 which is the scenario projecting the slowest pace of educational expansion, for three age groups (25–44, 45–64, 65+) and four reference years (2025, 2050, 2075, 2100). Even at the baseline year of 2025, a generalised female advantage in tertiary attainment is already observable across all age groups, with the bulk of the distribution sitting above the parity line (F/M = 1). This advantage is, however, less pronounced in the older age cohorts, reflecting the well-documented historical pattern whereby women entered higher education in large numbers only from the 1970s and 1980s onward, a cohort effect that progressively attenuates with age. Under SSP3, the female advantage is maintained throughout the projection horizon across all age groups. In the 25–44 cohort, the distribution shifts upward over time while simultaneously developing heavier tails, signalling an increasingly polarised cross-regional pattern in the sex dimension of educational expansion. A similar, if somewhat more moderate, dynamic is observed in the 45–64 group.

The 65+ age group exhibits a slightly different trajectory. The overall upward shift of the ratio is more pronounced, consistent with the well-established female advantage in life expectancy (particularly among the highly educated), which mechanically increases the share of women in the surviving highly educated 65+ population over time. Additionally, more pronounced distributional extremes emerge at longer projection horizons, a pattern that reflects both the amplification of small initial values in specific regions due to the length of the projection horizon and the compounding effects of female survival advantages and, in some contexts, selective female out-migration of the highly educated. These dynamics are partly inherent to the modelling framework when applied over very long horizons and should be interpreted with appropriate caution. Notwithstanding these nuances, and the non-negligible share of regions exhibiting female-to-male ratios at or below parity, the female advantage in tertiary educational attainment is sustained across the full projection horizon, albeit with varying magnitude and with an increasing degree of cross-regional heterogeneity as the projection period extends.

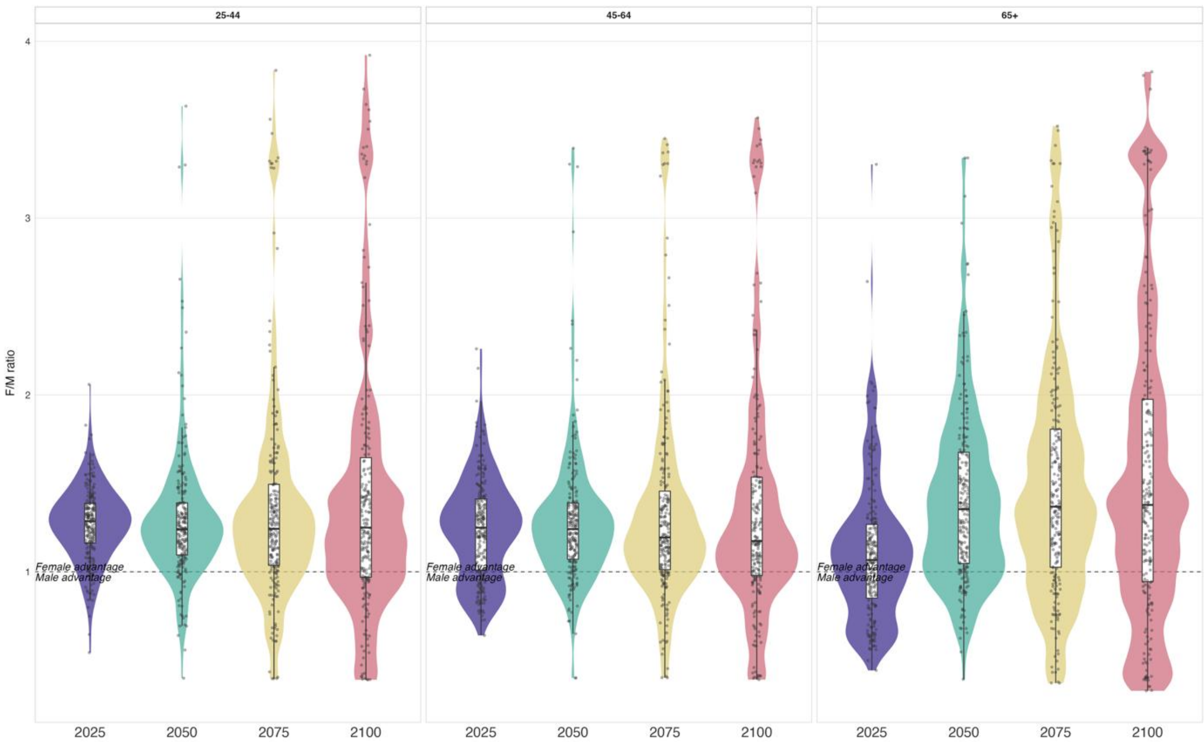

Figure 6: Distribution of the female to male ratio in tertiary educated population (ISCED 5-8) across NUTS-2 regions under SSP 3, for four selected years (2025, 2050, 2075, 2100) and three adult age groups (25 to 44, 45 to 64, and 65 and older). Each point represents a single NUTS-2 region. The dashed horizontal line marks parity (F/M = 1); values above indicate a female advantage in tertiary attainment, values below a male advantage. Violin shapes show the full distribution across regions; the embedded boxplot shows the median and interquartile range.

To further explore the sex dimension of educational expansion at the subnational level, Figure 7 examines the number of years required for each NUTS-2 region to achieve a 30-percentage point increase in the share of tertiary-educated population aged 25–44 relative to 2025 under SSP1. The 30-percentage point threshold was selected as it represents a substantial yet attainable expansion that the majority of regions cross within the projection horizon, making it informative for comparing the speed of educational expansion across regions. Two overarching patterns emerge from this visualisation. The first concerns the sex differential in the speed of educational expansion. Across virtually all regions and countries, female populations (circles) consistently reach the 30-percentage point threshold in fewer years than their male counterparts (triangles), as evidenced by the systematic leftward displacement of circles relative to triangles along the horizontal axis. The grey connecting lines make this gap immediately legible: in many regions it spans several decades, underscoring that the female advantage documented in the attainment levels translates into a structurally faster pace of expansion. The second pattern concerns the substantial within-country heterogeneity in the speed of educational expansion. Even under the most favourable scenario, the years required to reach the threshold vary considerably across NUTS-2 regions belonging to the same country, as reflected by the vertical spread of same-coloured symbols within each cluster. This dispersion is not merely a statistical artefact but reflects genuine differences in regional starting conditions, labour market structures, migration dynamics and institutional contexts that shape the pace at which educational expansion unfolds at the subnational level. Taken together, these two findings reinforce the analytical value of a subnational perspective: the regional and sex dimensions of educational expansion are not reducible to national averages and interact in ways that would remain invisible in country-level analyses.

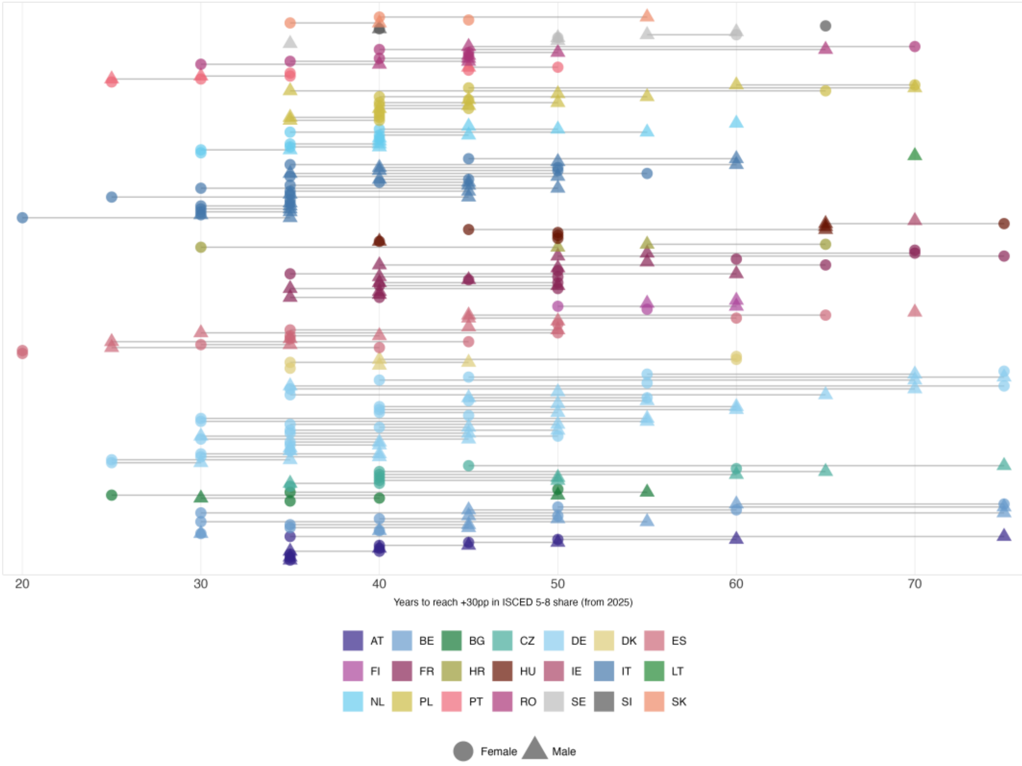

Figure 7: Number of years required for each NUTS-2 region to achieve a 30-percentage point increase in the share of the population aged 25 to 44 with tertiary education (ISCED 5-8), relative to 2025, under SSP 1. Each pair of symbols represents a single NUTS-2 region, with circles indicating females and triangles indicating males. The grey horizontal line connecting each pair shows the gap in speed between the two sexes for the same region. Regions are sorted within each country group by median speed. Regions for which the 30-percentage point threshold is not reached by 2100 are excluded. Colours identify countries.

When considered alongside the broader process of population ageing, the educational expansion documented above takes on a more nuanced character, as the absolute stock of highly educated young adults is shaped not only by attainment rates but also by the shrinking size of the 25–44 cohort itself.
The left panel of Figure 8 shows the projected percentage change in the overall size of this age group between 2025 and 2050, drawn from Tamburini et al. (2026), while the right panel presents the corresponding change in the absolute number of highly educated individuals (ISCED 5–8) within the same age group, as projected in this study. The juxtaposition reveals a fundamental tension at the heart of European human capital dynamics. Despite the broad and at times rapid expansion of educational attainment documented in the preceding sections, the concurrent process of population ageing exerts a countervailing force that, in many regions, matches or even exceeds the pace of educational growth in absolute terms. The spatial pattern that emerges over this 25-year horizon is strongly differentiated along two gradients. The first and most pronounced runs from east to west: across much of Central and Eastern Europe, the contraction of the 25–44 cohort driven by population ageing and sustained out-migration outpaces educational expansion, resulting in an absolute decline in the number of highly educated young adults despite improving attainment rates. The second, somewhat less pronounced gradient runs from south to north: the northwestern regions of Europe, where fertility has remained comparatively higher and immigration is projected to partially offset demographic decline, manage to sustain or even increase the absolute stock of highly skilled young adults, while many southern European regions struggle to compensate for the shrinkage of young cohorts through

educational gains alone. This result is analytically significant precisely because it would remain invisible in analyses conducted at either the national level or along a single demographic dimension. It underscores the fundamental motivation of this study: educational attainment and age structure are not independent processes, and their interaction at the subnational level produces outcomes that diverge sharply from what either trend would suggest in isolation.

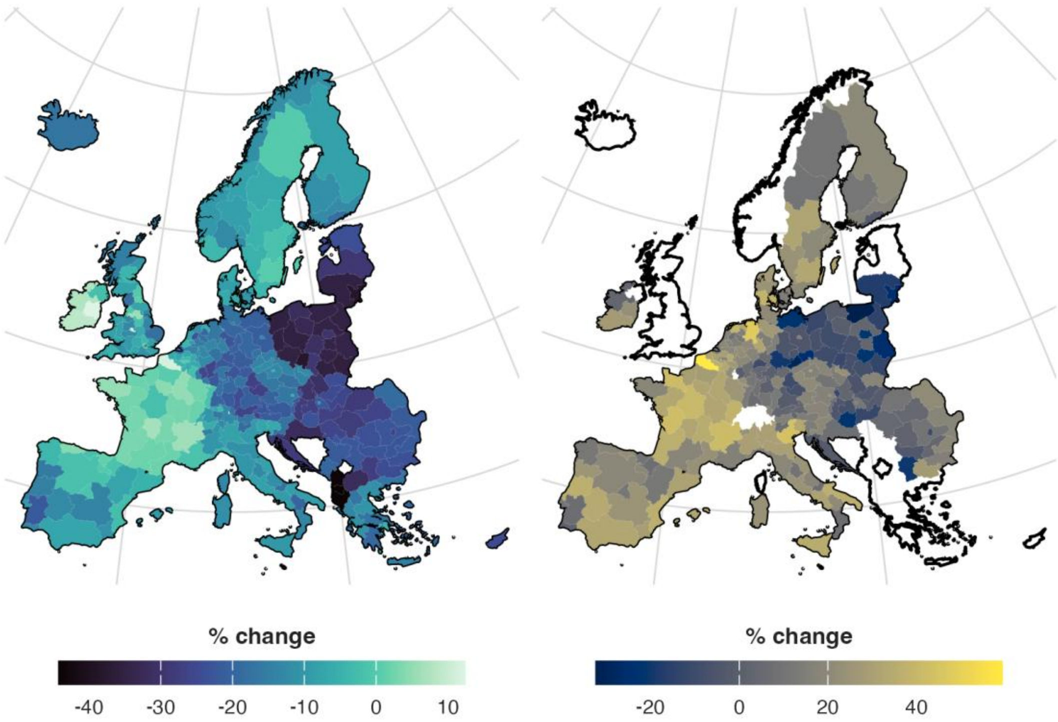

Figure 8: Percentage change between 2025 and 2050 in the size of the 25–44 age group (left panel) and in the absolute number of highly educated individuals (ISCED 5–8) aged 25–44 (right panel) across European NUTS-2 regions under SSP2. The left panel draws on the demographic projections of Tamburini et al. (2026), while the right panel presents results from the present study. Regions in white indicate missing data.

At the same time, the ageing of the European population documented in Tamburini et al. (2026) does not unfold in an educational vacuum. As shown in Figure 9, the 65 and above age group, which is the age group with the highest concentration of low educational attainment, undergoes a profound compositional transformation over the projection horizon (2025 to 2050 under SSP 2 as for the previous Figure 8). Comparing the upper and lower panels, the dominance of ISCED 0–2 that characterises the educational structure of this cohort in 2025 gives way, by 2075, to a distribution in which mid- and high-education categories account for a substantially larger and in many regions majority share. This shift is pervasive across countries and NUTS-2 regions, though its pace and magnitude vary considerably, reflecting the heterogeneous historical trajectories of educational expansion that are progressively feeding into the older age groups through cohort replacement. This transformation carries implications that extend well beyond the conventional framing of population ageing as a demographic burden. A more educated older population poses questions about potentially higher rates of labour force participation below statutory retirement ages, greater adaptability to technological change, and enhanced resilience to external shocks, including climate extremes. When considered alongside the absolute growth of this age group driven by sustained gains in life expectancy, the educational upgrading of the 65 and above cohort thus adds an dimension to the European human capital landscape.

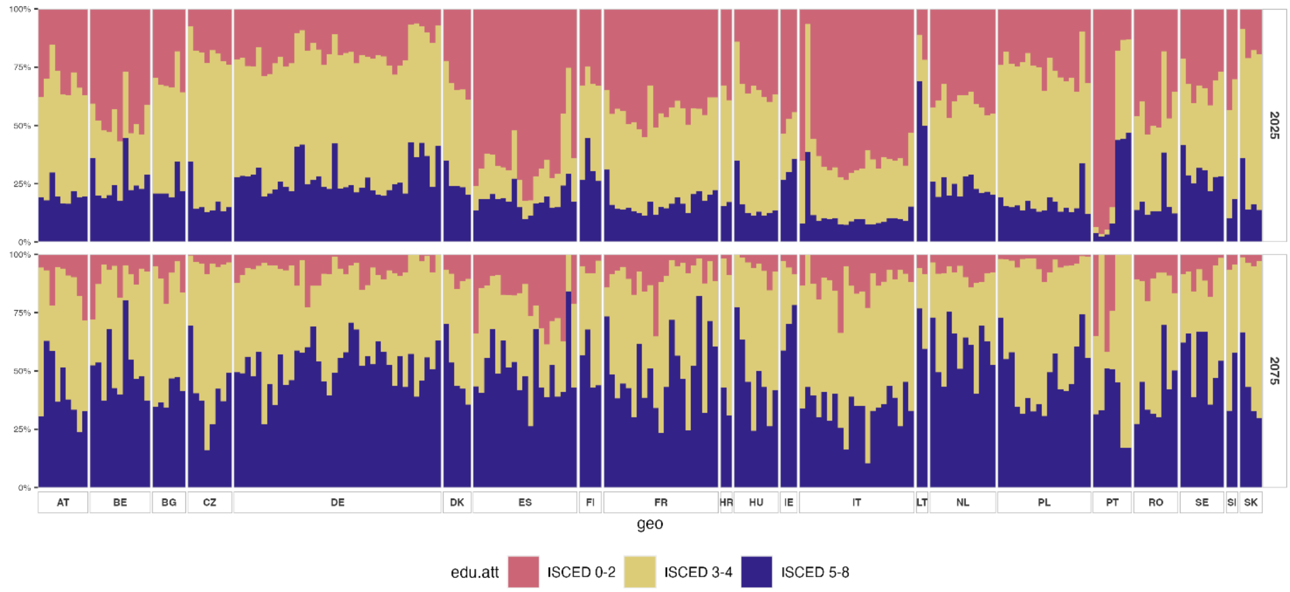

Figure 9: Educational composition of the population aged 65 and above across European NUTS-2 regions in 2025 (upper panel) and 2075 (lower panel) under SSP 2. Stacked bars show the share of each education level for each NUTS-2 region, grouped by country.

5 Conclusions

This study makes two main contributions to the literature on the formation and temporal evolution of human capital at the subnational level. The first is methodological: the framework employs Bayesian time series models, a class of methods distinguished by its natural suitability for hierarchical data structures and its capacity to explicitly quantify uncertainty in both model inputs and predictive outputs. The modelling architecture operates in the log-return space, exploiting the compositional nature of educational attainment distributions within subpopulations, and combines this with Iterative Proportional Fitting to enforce coherence with external marginal constraints and reduce residual uncertainty at the aggregate level. The second contribution is the resulting dataset. This constitutes an SSP-coherent projection of educational attainment, disaggregated into three ISCED group categories, for the age- and sex-specific populations of European NUTS-2 regions, covering three adult age groups from age 25 onward across 215 regions in 21 countries. To the best of the author's knowledge, this is the first publicly available dataset providing subnational human capital projections by age, sex, and educational attainment under the full SSP framework at this scale of European regional coverage. The projection results align with, and in several respects even extend, existing strands of the literature. The spatially pervasive yet heterogeneous expansion of tertiary attainment documented across European NUTS-2 regions confirms and enriches the convergence dynamics analysed by Dańska-Borsiak (2023). The core-periphery educational disparities identified by Hippe (2020) remain visible well until near the end of the projection horizon, though their character evolves: in some regions, the broad expansion of tertiary attainment progressively compresses the historical advantage of leading areas, while in others the gap persists, likely reflecting the continued clustering of human capital in large metropolitan centres. A further consistent trend, interacting structurally with age composition, is the female advantage in tertiary attainment, which is observed and confirmed across all age groups and throughout the full projection horizon suggesting the continuation of the ongoing reversal in sex-specific educational advantage (Grow and Van Bavel 2015; Corti and Scherer 2022). The results also shed additional light on the critical interaction between educational expansion and the Europe-wide process of population ageing, itself characterised by substantial subnational variation. On one hand, smaller younger cohorts are capable of partially offsetting, in terms of absolute stocks rather than average attainment rates, the gains generated by educational expansion. This is a tension that is spatially concentrated in Central and Eastern European regions. On the other hand, the progressive entry of increasingly educated cohorts into the 65 and above age group produces a population that is simultaneously larger and better educated, reframing the conventional narrative of demographic ageing as an unambiguous erosion of the human capital stock.

The present study addresses a problem that is inherently challenging from both a data availability and a methodological standpoint. Several limitations of the present framework should therefore be acknowledged explicitly. The first concerns the absence of a cohort-component structure. Educational attainment is modelled here as a stochastic process informed by age and sex dimensions, but without explicit educational transition rates or cohort replacement mechanics. This departs from the multi-state tradition established by Lutz and colleagues and means that the intergenerational transmission of education is not directly represented in the model in a cohort fashion. This is, of course, also due to the reliance on the age, sex, and region-specific projections from Tamburini et al. (2026), which are also obtained outside of a cohort model context and with age group sizes and projection steps that do not allow for a usual cohort approach. A related limitation is that these broad chronological age groups may only imperfectly capture individuals' functional capacities, labour-market potential, health status, or social roles, particularly at older ages. Second, the model is estimated on a maximum of twenty years of annual regional observations, which constitutes a thin empirical basis for projections extending to 2100. While anchoring to national SSP trajectories and SSP-coherent regional age structures partially compensate for this, uncertainty compounds substantially over the projection horizon. This limitation is compounded by the sampling variability inherent in the Labour Force Survey data used for estimation, which is particularly pronounced in smaller regions and specific age-education cells and introduces noise into the historical series from which trend parameters are derived. Third, scenario differentiation across SSPs is externally imposed rather than endogenously modelled: regional educational trajectories are not projected from first principles under each scenario but instead reflect the national SSP narratives propagated downward through the IPF adjustment and the anchoring to SSP-coherent age structures. The subnational scenario spread is therefore an inheritance of national-level assumptions rather than the product of genuinely regionalised scenario logic. Finally, a residual tension persists between the Eurostat data used as the historical basis and the WCDE national projections. The decision to standardise to Eurostat values for the historical period is methodologically defensible, but it introduces some ambiguity in how the SSP narratives map onto the regional starting conditions from which the projections depart. Notwithstanding these limitations, the dataset opens avenues for a richer exploration of the socioeconomic consequences of human capital development, including its interactions with labour force participation, fertility, productivity, and the capacity for climate change adaptation. Spatially explicit human capital projections of this kind can serve as a key interface between demographic research, economic foresight, and climate resilience planning, providing a granular empirical basis for evidence-informed strategies tailored to Europe's diverse regional trajectories.

Several directions for future work follow naturally from the limitations identified above. On the methodological side, priority should be given to developing a technically enhanced framework capable of jointly representing age, sex, and education trajectories within the SSP narratives, drawing on a broader array of harmonised data sources. These could include national micro-censuses and educational transition statistics to better constrain regional uncertainties. On the spatial side, further refinement is needed to capture intra-regional heterogeneity and cross-border functional linkages that current NUTS-2 boundaries obscure: many NUTS-2 regions are large enough to encompass structurally distinct urban and rural areas whose dynamics are averaged away at this level of aggregation, making an extension to NUTS-3 and LAU geographies a natural next step. Finally, the development of genuinely regionalised SSP narratives would represent a substantive advance, enabling subnational projections that reflect region-specific demographic and institutional dynamics rather than inheriting them wholesale from the national level.

References

Aitchison, J. 1986. *The Statistical Analysis of Compositional Data*. Springer Netherlands. https://doi.org/10.1007/978-94-009-4109-0.

Alkema, Leontine, Adrian E. Raftery, Patrick Gerland, et al. 2011. 'Probabilistic Projections of the Total Fertility Rate for All Countries'. *Demography* 48 (3): 815–39. https://doi.org/10.1007/s13524-011-0040-5.

Barro, Robert J., and Jong Wha Lee. 2013. 'A New Data Set of Educational Attainment in the World, 1950–2010'. *Journal of Development Economics* 104 (September): 184–98. https://doi.org/10.1016/j.jdeveco.2012.10.001.

Becker, Gary S. 1993. *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. University of Chicago Press. https://doi.org/10.7208/chicago/9780226041223.001.0001.

Betthäuser, Bastian A., Caspar Kaiser, and Nhat An Trinh. 2021. 'Regional Variation in Inequality of Educational Opportunity across Europe'. *Socius: Sociological Research for a Dynamic World* 7 (January): 23780231211019890. https://doi.org/10.1177/23780231211019890.

Bijak, Jakub, and John Bryant. 2016. 'Bayesian Demography 250 Years after Bayes'. *Population Studies* 70 (1): 1–19. https://doi.org/10.1080/00324728.2015.1122826.

Box, G. E. P., and D. R. Cox. 1964. 'An Analysis of Transformations'. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 26 (2): 211–43. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x.

Brunello, Giorgio, Margherita Fort, Nicole Schneeweis, and Rudolf Winter-Ebmer. 2016. 'The Causal Effect of Education on Health: What Is the Role of Health Behaviors?' *Health Economics* 25 (3): 314–36. https://doi.org/10.1002/hec.3141.

Bryant, John, and Junni L. Zhang. 2018. *Bayesian Demographic Estimation and Forecasting*. 1st edn. Chapman and Hall/CRC. https://doi.org/10.1201/9780429452987.

Calvin, Katherine, Ben Bond-Lamberty, Leon Clarke, et al. 2017. 'The SSP4: A World of Deepening Inequality'. *Global Environmental Change* 42 (January): 284–96. https://doi.org/10.1016/j.gloenvcha.2016.06.010.

Card, David. 1999. 'The Causal Effect of Education on Earnings'. In *Handbook of Labor Economics*, vol. 3. Elsevier. https://doi.org/10.1016/S1573-4463(99)03011-4.

Castelló-Climent, Amparo, and Rafael Domenech. 2022. 'Converging to Convergence: The Role of Human Capital'. Preprint, In Review, December 6. https://doi.org/10.21203/rs.3.rs-2335045/v1.

Chao, Fengqing, Patrick Gerland, Alex R. Cook, and Leontine Alkema. 2021. 'Global Estimation and Scenario-Based Projections of Sex Ratio at Birth and Missing Female Births Using a Bayesian Hierarchical Time Series Mixture Model'. *The Annals of Applied Statistics* 15 (3). https://doi.org/10.1214/20-AOAS1436.

Chocholatá, Michaela, and Andrea Furková. 2017. 'Regional Disparities in Education Attainment Level in the European Union: A Spatial Approach'. *Baltic Journal of European Studies* 7 (2): 107–31. https://doi.org/10.1515/bjes-2017-0012.

Corti, Giulia, and Stefani Scherer. 2022. 'Find the Right One. Educational Assortative Mating and Educational Reproduction in Germany'. *Research in Social Stratification and Mobility* 81 (October): 100716. https://doi.org/10.1016/j.rssm.2022.100716.

Curtale, Riccardo, Matteo Schiavone, and Filipe Batista e Silva. 2026. 'A Novel Approach to Regionalize Country-Level GDP Projections'. *Journal of Forecasting* 45 (2): 867–79. https://doi.org/10.1002/for.70052.

Cutler, David M., and Adriana Lleras-Muney. 2010. 'Understanding Differences in Health Behaviors by Education'. *Journal of Health Economics* 29 (1): 1–28. https://doi.org/10.1016/j.jhealeco.2009.10.003.

Dańska-Borsiak, Barbara. 2023. 'Human Capital Convergence in European NUTS 2 Regions'. *Equilibrium. Quarterly Journal of Economics and Economic Policy* 18 (2): 367–92. https://doi.org/10.24136/eq.2023.011.

Dee, Thomas S. 2004. 'Are There Civic Returns to Education?' *Journal of Public Economics* 88 (9–10): 1697–720. https://doi.org/10.1016/j.jpubeco.2003.11.002.

Diebolt, Claude, and Ralph Hippe. 2017. 'Regional Human Capital Inequality in Europe in the Long Run, 1850-2010'. *Région et Développement* 45 (October): 5–30.
Diebolt, Claude, and Ralph Hippe. 2018. 'Remoteness Equals Backwardness? Human Capital and Market Access in the European Regions: Insights from the Long Run'. *Education Economics* 26 (3): 285–304. https://doi.org/10.1080/09645292.2017.1401979.
Diebolt, Claude, and Ralph Hippe. 2019. 'The Long-Run Impact of Human Capital on Innovation and Economic Development in the Regions of Europe'. *Applied Economics* 51 (5): 542–63. https://doi.org/10.1080/00036846.2018.1495820.
Duran, Dan Cristian, Luminita Maria Gogan, Alin Artene, and Vasile Duran. 2015. 'The Components of Sustainable Development - A Possible Approach'. *Procedia Economics and Finance* 26: 806–11. https://doi.org/10.1016/S2212-5671(15)00849-7.
Égert, Balázs, Jarmila Botev, and David Turner. 2020. 'The Contribution of Human Capital and Its Policies to per Capita Income in Europe and the OECD'. *European Economic Review* 129 (October): 103560. https://doi.org/10.1016/j.euroecorev.2020.103560.
Egozcue, J. J., V. Pawlowsky-Glahn, G. Mateu-Figueras, and C. Barceló-Vidal. 2003. 'Isometric Logratio Transformations for Compositional Data Analysis'. *Mathematical Geology* 35 (3): 279–300. https://doi.org/10.1023/A:1023818214614.
Emelyanova, Anastasia. 2019. 'Exploring the Future Population and Educational Dynamics in the Arctic: 2015 to 2050'. *Finnish Yearbook of Population Research* 53 (September): 1–24. https://doi.org/10.23979/fypr.70159.
Eurofound (European Foundation for the Improvement of Living and Working Conditions). 2024. *Role of Human Capital Inequalities in Social Cohesion and Convergence.* Publications Office. https://doi.org/10.2806/143041.
European Commission. Joint Research Centre. 2018. *Demographic and Human Capital Scenarios for the 21st Century :2018 Assessment for 201 Countries.* Publications Office. https://doi.org/10.2760/41776.
Eurostat. 2024a. 'Population on 1st January by Age, Sex and Type of Projection'. Eurostat. https://doi.org/10.2908/PROJ_23NP.
Eurostat. 2024b. 'Population on 1st January by Age, Sex and Type of Projection'. Eurostat. https://doi.org/10.2908/PROJ_23NP.
Eurostat. 2025a. 'Bevölkerung in Privathaushalten Nach Bildungsabschluss Und NUTS-2-Region'. Eurostat. https://doi.org/10.2908/LFST_R_LFSD2POP.
Eurostat. 2025b. 'Life Expectancy at Birth by Sex'. Eurostat. https://doi.org/10.2908/TPS00208.
Eurostat. 2025c. 'Total Fertility Rate'. Eurostat. https://doi.org/10.2908/TPS00199.
Fienberg, Stephen E. 1970. 'An Iterative Procedure for Estimation in Contingency Tables'. *The Annals of Mathematical Statistics* 41 (3): 907–17. https://doi.org/10.1214/aoms/1177696968.
Fujimori, Shinichiro, Tomoko Hasegawa, Toshihiko Masui, et al. 2017. 'SSP3: AIM Implementation of Shared Socioeconomic Pathways'. *Global Environmental Change* 42 (January): 268–83. https://doi.org/10.1016/j.gloenvcha.2016.06.009.
Galor, Oded, and Daniel Tsiddon. 1997. 'The Distribution of Human Capital and Economic Growth'. *Journal of Economic Growth* 2 (1): 93–124. https://doi.org/10.1023/A:1009785714248.
Gelman, Andrew, and Donald B. Rubin. 1992. 'Inference from Iterative Simulation Using Multiple Sequences'. *Statistical Science* 7 (4). https://doi.org/10.1214/ss/1177011136.
Goujon, A, and K. McNay. 2003. 'Projecting the Educational Composition of the Population of India: Selected State-Level Perspectives'. *Applied Population and Policy* 1 (1): 25–35.
Goujon, Anne, Samir KC, Markus Speringer, et al. 2016. 'A Harmonized Dataset on Global Educational Attainment between 1970 and 2060 - An Analytical Window into Recent Trends and Future Prospects in Human Capital Development'. *Journal of Demographic Economics* 82 (3): 315–63. https://doi.org/10.1017/dem.2016.10.
Goujon, Anne, Iliana Kohler, and W. Lutz. 2000. *Future Population and Education Trends: Scenarios to 2030 by Socioecological Region*.
Grow, André, and Jan Van Bavel. 2015. 'Assortative Mating and the Reversal of Gender Inequality in Education in Europe: An Agent-Based Model'. *PLOS ONE* 10 (6): e0127806. https://doi.org/10.1371/journal.pone.0127806.
Hanushek, Eric A., and Ludger Woessmann. 2008. 'The Role of Cognitive Skills in Economic Development'. *Journal of Economic Literature* 46 (3): 607–68. https://doi.org/10.1257/jel.46.3.607.

Heckman, James J., John Eric Humphries, and Gregory Veramendi. 2018. 'Returns to Education: The Causal Effects of Education on Earnings, Health, and Smoking'. *Journal of Political Economy* 126 (S1): S197–246. https://doi.org/10.1086/698760.

Hippe, R. 2013. 'Spatial Clustering of Human Capital in the European Regions'. *Economies et Sociétés (Serie 'Histoire Economique Quantitative')* None (46): 1077–104. https://doi.org/None.

Hippe, Ralph. 2020. 'Human Capital in European Regions since the French Revolution: Lessons for Economic and Education Policies': *Revue d'économie Politique* Vol. 130 (1): 27–50. https://doi.org/10.3917/redp.301.0027.

Hippe, Ralph, and Joerg Baten. 2012. 'Regional Inequality in Human Capital Formation in Europe, 1790–1880'. *Scandinavian Economic History Review* 60 (3): 254–89. https://doi.org/10.1080/03585522.2012.727763.

Iammarino, Simona, Andrés Rodriguez-Pose, and Michael Storper. 2019. 'Regional Inequality in Europe: Evidence, Theory and Policy Implications'. *Journal of Economic Geography* 19 (2): 273–98. https://doi.org/10.1093/jeg/lby021.

Irwin, V., T. M. Bailey, R. Panditharatna, and A. Sadeghi. 2024. *Projections of Education Statistics to 2030*. NCES 2024-034. National Center for Education Statistics. 2024034. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2024034.

James, Friday, and Joshua Weese. 2022. 'Neural Network-Based Forecasting of Student Enrollment With Exponential Smoothing Baseline and Performance Analysis'. *2022 ASEE Annual Conference & Exposition Proceedings*, August, 41751. https://doi.org/10.18260/1-2--41751.

Kc, Samir, Bilal Barakat, Anne Goujon, Vegard Skirbekk, Warren C. Sanderson, and Wolfgang Lutz. 2010. 'Projection of Populations by Level of Educational Attainment, Age, and Sex for 120 Countries for 2005-2050'. *Demographic Research* 22 (March): 383–472. https://doi.org/10.4054/DemRes.2010.22.15.

KC, Samir, Mohit Dhakad, Michaela Potančoková, et al. 2024. *Updating the Shared Socioeconomic Pathways (SSPs) Global Population and Human Capital Projections*. WP-24-003. International Institute for Applied Systems Analysis (IIASA). https://pure.iiasa.ac.at/.

Kc, Samir, and Wolfgang Lutz. 2017. 'The Human Core of the Shared Socioeconomic Pathways: Population Scenarios by Age, Sex and Level of Education for All Countries to 2100'. *Global Environmental Change* 42 (January): 181–92. https://doi.org/10.1016/j.gloenvcha.2014.06.004.

Kc, Samir, Marcus Wurzer, Markus Speringer, and Wolfgang Lutz. 2018. 'Future Population and Human Capital in Heterogeneous India'. *Proceedings of the National Academy of Sciences* 115 (33): 8328–33. https://doi.org/10.1073/pnas.1722359115.

Keilman, Nico. 2020. 'Uncertainty in Population Forecasts for the Twenty-First Century'. *Annual Review of Resource Economics* 12 (1): 449–70. https://doi.org/10.1146/annurev-resource-110319-114841.

Kotschy, Rainer, Patricio Suarez Urtaza, and Uwe Sunde. 2020. 'The Demographic Dividend Is More than an Education Dividend'. *Proceedings of the National Academy of Sciences* 117 (42): 25982–84. https://doi.org/10.1073/pnas.2012286117.

Kraay, Aart. 2018. *Methodology for a World Bank Human Capital Index*. Policy Research Working Paper No. 8593. The World Bank. https://econpapers.repec.org/RePEc:wbk:wbrwps:8593.

Kynčlová, Petra, Peter Filzmoser, and Karel Hron. 2015. 'Modeling Compositional Time Series with Vector Autoregressive Models'. *Journal of Forecasting* 34 (4): 303–14. https://doi.org/10.1002/for.2336.

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

Lutz, Wolfgang. 2009. '*Sola Schola et Sanitate*: Human Capital as the Root Cause and Priority for International Development?' *Philosophical Transactions of the Royal Society B: Biological Sciences* 364 (1532): 3031–47. https://doi.org/10.1098/rstb.2009.0156.

Lutz, Wolfgang. 2013. 'Demographic Metabolism: A Predictive Theory of Socioeconomic Change'. *Population and Development Review* 38 (s1): 283–301. https://doi.org/10.1111/j.1728-4457.2013.00564.x.

Lutz, Wolfgang. 2017. 'Global Sustainable Development Priorities 500 y after Luther: *Sola Schola et Sanitate*'. *Proceedings of the National Academy of Sciences* 114 (27): 6904–13. https://doi.org/10.1073/pnas.1702609114.

Lutz, Wolfgang, William P. Butz, and Samir Kc, eds. 2014. *World Population and Human Capital in the Twenty-First Century*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198703167.001.0001.

Lutz, Wolfgang, Jesus Crespo Cuaresma, Endale Kebede, Alexia Prskawetz, Warren C. Sanderson, and Erich Striessnig. 2019. 'Education Rather than Age Structure Brings Demographic Dividend'. *Proceedings of the National Academy of Sciences* 116 (26): 12798–803. https://doi.org/10.1073/pnas.1820362116.
Lutz, Wolfgang, Jesus Crespo Cuaresma, and Warren Sanderson. 2008. 'The Demography of Educational Attainment and Economic Growth'. *Science* 319 (5866): 1047–48. https://doi.org/10.1126/science.1151753.
Lutz, Wolfgang, and Anne Goujon. 2001a. 'The World's Changing Human Capital Stock: Multi-State Population Projections by Educational Attainment'. *Population and Development Review* 27 (2): 323–39. https://doi.org/10.1111/j.1728-4457.2001.00323.x.
Lutz, Wolfgang, and Anne Goujon. 2001b. 'The World's Changing Human Capital Stock: Multi-State Population Projections by Educational Attainment'. *Population and Development Review* 27 (2): 323–39. https://doi.org/10.1111/j.1728-4457.2001.00323.x.
Lutz Wolfgang, Goujon Anne, K.C. Samir, Stonawski Marcin, and Stilianakis Nikolaos. 2018. *Demographic and Human Capital Scenarios for the 21st Century: 2018 Assessment for 201 Countries.* Publications Office. https://data.europa.eu/doi/10.2760/835878.
Lutz, Wolfgang, Anne Goujon, and Annababette Wils. 2008. 'The Population Dynamics of Human Capital Accumulation'. *Population and Development Review* 34: 149–87. JSTOR.
Lutz, Wolfgang, and Samir Kc. 2011. 'Global Human Capital: Integrating Education and Population'. *Science* 333 (6042): 587–92. https://doi.org/10.1126/science.1206964.
Lutz, Wolfgang, Raya Muttarak, and Erich Striessnig. 2014. 'Universal Education Is Key to Enhanced Climate Adaptation'. *Science* 346 (6213): 1061–62. https://doi.org/10.1126/science.1257975.
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.
Mao, Shengzhong, Chaoli Zhang, Yichi Song, et al. 2024. 'Time Series Analysis for Education: Methods, Applications, and Future Directions'. arXiv:2408.13960. Preprint, arXiv, August 27. https://doi.org/10.48550/arXiv.2408.13960.
Marois, Guillaume, Alain Bélanger, and Wolfgang Lutz. 2020. 'Population Aging, Migration, and Productivity in Europe'. *Proceedings of the National Academy of Sciences* 117 (14): 7690–95. https://doi.org/10.1073/pnas.1918988117.
Marois, Guillaume, Michaela Potančoková, Agnieszka Bezat, and Jesús Crespo Cuaresma. 2026. 'Projecting Labour Market Imbalances and Skill Mismatch Under Demographic Change in the EU'. *European Journal of Population* 42 (1): 4. https://doi.org/10.1007/s10680-025-09758-2.
Marois, Guillaume, Patrick Sabourin, and Alain Bélanger. 2019. 'How Reducing Differentials in Education and Labor Force Participation Could Lessen Workforce Decline in the EU-28'. *Demographic Research* 41 (July): 125–60. https://doi.org/10.4054/DemRes.2019.41.6.
Mincer, Jacob. 1958. 'Investment in Human Capital and Personal Income Distribution'. *Journal of Political Economy* 66 (4): 281–302. https://doi.org/10.1086/258055.
OECD. 2001. *The Well-Being of Nations: The Role of Human and Social Capital*. OECD. https://doi.org/10.1787/9789264189515-en.
OECD. 2023. *Education at a Glance 2023: OECD Indicators*. Education at a Glance. OECD Publishing. https://doi.org/10.1787/e13bef63-en.
O'Neill, Brian C., Elmar Kriegler, Kristie L. Ebi, et al. 2017. 'The Roads Ahead: Narratives for Shared Socioeconomic Pathways Describing World Futures in the 21st Century'. *Global Environmental Change* 42 (January): 169–80. https://doi.org/10.1016/j.gloenvcha.2015.01.004.
O'Neill, Brian C., Elmar Kriegler, Keywan Riahi, et al. 2014. 'A New Scenario Framework for Climate Change Research: The Concept of Shared Socioeconomic Pathways'. *Climatic Change* 122 (3): 387–400. https://doi.org/10.1007/s10584-013-0905-2.
Pelinescu, Elena. 2015. 'The Impact of Human Capital on Economic Growth'. *Procedia Economics and Finance* 22: 184–90. https://doi.org/10.1016/S2212-5671(15)00258-0.
Pérez, José Ramón. 2026. 'Brain Drain: How the Exodus of Talent Is Redefining the Map of Europe'. *European Data Journalism Network (EDJNet)*, February 12. https://www.europeandatajournalism.eu/cp_data_news/brain-drain-how-the-exodus-of-talent-is-redefining-the-map-of-europe/.
Persson, Mikael. 2014. 'Testing the Relationship Between Education and Political Participation Using the 1970 British Cohort Study'. *Political Behavior* 36 (4): 877–97. https://doi.org/10.1007/s11109-013-9254-0.

Plummer, Martyn. 2003. ‘JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling’. *3rd International Workshop on Distributed Statistical Computing (DSC 2003); Vienna, Austria* 124 (April).
Potančoková, Michaela, Guillaume Marois, and Miguel Gonzalez Leonardo. 2023. ‘QuantMig Microsimulation Population Projection Model and Migration Scenarios for 31 European Countries’. Version v1.0. With Jakub Bijak. Zenodo, August 4. https://doi.org/10.5281/ZENODO.7728049.
Prenzel, Paula, and Simona Iammarino. 2021. ‘Labor Force Aging and the Composition of Regional Human Capital’. *Economic Geography* 97 (2): 140–63. https://doi.org/10.1080/00130095.2021.1885294.
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.
R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, released. https://www.R-project.org/.
Raftery, Adrian E., Leontine Alkema, and Patrick Gerland. 2014. ‘Bayesian Population Projections for the United Nations’. *Statistical Science* 29 (1). https://doi.org/10.1214/13-STS419.
Rao, Narasimha D., Petra Sauer, Matthew Gidden, and Keywan Riahi. 2019. ‘Income Inequality Projections for the Shared Socioeconomic Pathways (SSPs)’. *Futures* 105 (January): 27–39. https://doi.org/10.1016/j.futures.2018.07.001.
Reid, Colleen E., Marie S. O’Neill, Carina J. Gronlund, et al. 2009. ‘Mapping Community Determinants of Heat Vulnerability’. *Environmental Health Perspectives* 117 (11): 1730–36. https://doi.org/10.1289/ehp.0900683.
Riahi, Keywan, Detlef P. Van Vuuren, Elmar Kriegler, et al. 2017. ‘The Shared Socioeconomic Pathways and Their Energy, Land Use, and Greenhouse Gas Emissions Implications: An Overview’. *Global Environmental Change* 42 (January): 153–68. https://doi.org/10.1016/j.gloenvcha.2016.05.009.
Rodríguez-Pose, Andrés, and Vassilis Tselios. 2009. ‘EDUCATION AND INCOME INEQUALITY IN THE REGIONS OF THE EUROPEAN UNION*’. *Journal of Regional Science* 49 (3): 411–37. https://doi.org/10.1111/j.1467-9787.2008.00602.x.
Rodríguez-Pose, Andrés, and Vassilis Tselios. 2010. ‘Inequalities in Income and Education and Regional Economic Growth in Western Europe’. *The Annals of Regional Science* 44 (2): 349–75. https://doi.org/10.1007/s00168-008-0267-2.
Rodríguez-Pose, Andrés, and Vassilis Tselios. 2011. ‘Mapping the European Regional Educational Distribution’. *European Urban and Regional Studies* 18 (4): 358–74. https://doi.org/10.1177/0969776411399345.
Romer, Paul. 1989. *Human Capital And Growth: Theory and Evidence*. No. W3173. National Bureau of Economic Research. https://doi.org/10.3386/w3173.
Ryder, Norman B. 1965. ‘The Cohort as a Concept in the Study of Social Change’. *American Sociological Review* 30 (6): 843. https://doi.org/10.2307/2090964.
S. K.C., M. Speringer, M. Wurzer, and Wolfgang Lutz, Program Director World Population Program. 2017. ‘Population Projection by Age, Sex, and Educational Attainment in Rural and Urban Regions of 35 Provinces of India, 2011-2101: Technical Report on Projecting the Regionally Explicit Socioeconomic Heterogeneity in India’. In *Population Projection by Age, Sex, and Educational Attainment in Rural and Urban Regions of 35 Provinces of India, 2011-2101: Technical Report on Projecting the Regionally Explicit Socioeconomic Heterogeneity in India*. WP-17-004, April 3.
Sestito, B., L. Reimann, M. Mazzoleni, W. J. W. Botzen, and J. C. J. H. Aerts. 2025. ‘Identifying Vulnerability Factors Associated with Heatwave Mortality: A Spatial Statistical Analysis across Europe’. *Environmental Research Letters* 20 (4): 044025. https://doi.org/10.1088/1748-9326/adbcc8.
Sinuany-Stern, Zilla. 2021. ‘Forecasting Methods in Higher Education: An Overview’. In *Handbook of Operations Research and Management Science in Higher Education*, edited by Zilla Sinuany-Stern, vol. 309. International Series in Operations Research & Management Science. Springer International Publishing. https://doi.org/10.1007/978-3-030-74051-1_5.
Striessnig, Erich, Wolfgang Lutz, and Anthony G. Patt. 2013. ‘Effects of Educational Attainment on Climate Risk Vulnerability’. *Ecology and Society* 18 (1): art16. https://doi.org/10.5751/ES-05252-180116.
Tamburini, A., C. Bosco, and S. Striessnig. 2025. *A Neural Network Architecture for Disaggregating Age-Specific Population Projections to the Sub-National Level*. WP-25-003. International Institute for Applied Systems Analysis (IIASA). https://pure.iiasa.ac.at/id/eprint/20889.
Tamburini, Andrea, Claudio Bosco, and Erich Striessnig. 2026. ‘NUTS-2 Level SSP-Coherent Population Projections by Age and Sex.’ Version 1.0.0. Zenodo, January 15. https://doi.org/10.5281/ZENODO.18256234.

Tang, Hui-Wen Vivian, and Tzu-chin Rojoice Chou. 2016. 'On the Fit and Forecasting Performance of Grey Prediction Models for Projecting Educational Attainment'. *Kybernetes* 45 (9): 1387–405. https://doi.org/10.1108/K-03-2014-0050.

Tang, Hui-Wen Vivian, and Mu-Shang Yin. 2012. 'Forecasting Performance of Grey Prediction for Education Expenditure and School Enrollment'. *Economics of Education Review* 31 (4): 452–62. https://doi.org/10.1016/j.econedurev.2011.12.007.

Terama, Emma, Elizabeth Clarke, Mark D. A. Rounsevell, Stefan Fronzek, and Timothy R. Carter. 2019. 'Modelling Population Structure in the Context of Urban Land Use Change in Europe'. *Regional Environmental Change* 19 (3): 667–77. https://doi.org/10.1007/s10113-017-1194-5.

Tsay, Ruey S. 2010. *Analysis of Financial Time Series*. 1st edn. Wiley Series in Probability and Statistics. Wiley. https://doi.org/10.1002/9780470644560.

UNDP (United Nations Development Programme). 2025. 'Human Development Report 2025'. *UNDP (United Nations Development Programme)* (New York). https://report.hdr.undp.org.

Wiśniowski, Arkadiusz, Peter W. F. Smith, Jakub Bijak, James Raymer, and Jonathan J. Forster. 2015. 'Bayesian Population Forecasting: Extending the Lee-Carter Method'. *Demography* 52 (3): 1035–59. https://doi.org/10.1007/s13524-015-0389-y.

Yang, Stephanie, Hsueh-Chih Chen, Wen-Ching Chen, and Cheng-Hong Yang. 2020. 'Student Enrollment and Teacher Statistics Forecasting Based on Time-Series Analysis'. *Computational Intelligence and Neuroscience* 2020 (September): 1–15. https://doi.org/10.1155/2020/1246920.

Yang, Xiaoxuan. 2020. 'Health Expenditure, Human Capital, and Economic Growth: An Empirical Study of Developing Countries'. *International Journal of Health Economics and Management* 20 (2): 163–76. https://doi.org/10.1007/s10754-019-09275-w.

Zhu, Shujin, and Renyu Li. 2017. 'Economic Complexity, Human Capital and Economic Growth: Empirical Research Based on Cross-Country Panel Data'. *Applied Economics* 49 (38): 3815–28. https://doi.org/10.1080/00036846.2016.1270413.

Appendix

Country	N_Regions	Region_Codes
AT	9	AT11, AT12, AT13, AT21, AT22, AT31, AT32, AT33, AT34
BE	11	BE10, BE21, BE22, BE23, BE24, BE25, BE31, BE32, BE33, BE34, BE35
BG	6	BG31, BG32, BG33, BG34, BG41, BG42
CZ	8	CZ01, CZ02, CZ03, CZ04, CZ05, CZ06, CZ07, CZ08
DE	38	DE11, DE12, DE13, DE14, DE21, DE22, DE23, DE24, DE25, DE26, DE27, DE30, DE40, DE50, DE60, DE71, DE72, DE73, DE80, DE91, DE92, DE93, DE94, DEA1, DEA2, DEA3, DEA4, DEA5, DEB1, DEB2, DEB3, DEC0, DED2, DED4, DED5, DEE0, DEF0, DEG0
DK	5	DK01, DK02, DK03, DK04, DK05
ES	19	ES11, ES12, ES13, ES21, ES22, ES23, ES24, ES30, ES41, ES42, ES43, ES51, ES52, ES53, ES61, ES62, ES63, ES64, ES70
FI	4	FI19, FI1B, FI1C, FI1D
FR	21	FR10, FRB0, FRC1, FRC2, FRD1, FRD2, FRE1, FRE2, FRF1, FRF2, FRF3, FRG0, FRH0, FRI1, FRI2, FRI3, FRJ1, FRJ2, FRK1, FRK2, FRL0
HR	2	HR03, HR04
HU	8	HU11, HU12, HU21, HU22, HU23, HU31, HU32, HU33
IE	3	IE04, IE05, IE06
IT	21	ITC1, ITC2, ITC3, ITC4, ITF1, ITF2, ITF3, ITF4, ITF5, ITF6, ITG1, ITG2, ITH1, ITH2, ITH3, ITH4, ITH5, ITI1, ITI2, ITI3, ITI4
LT	2	LT01, LT02
NL	12	NL11, NL12, NL13, NL21, NL22, NL23, NL31, NL32, NL33, NL34, NL41, NL42
PL	17	PL21, PL22, PL41, PL42, PL43, PL51, PL52, PL61, PL62, PL63, PL71, PL72, PL81, PL82, PL84, PL91, PL92
PT	7	PT11, PT15, PT16, PT17, PT18, PT20, PT30
RO	8	RO11, RO12, RO21, RO22, RO31, RO32, RO41, RO42
SE	8	SE11, SE12, SE21, SE22, SE23, SE31, SE32, SE33
SI	2	SI03, SI04
SK	4	SK01, SK02, SK03, SK04

Table A1: Age-sex-education dimension geographical scope

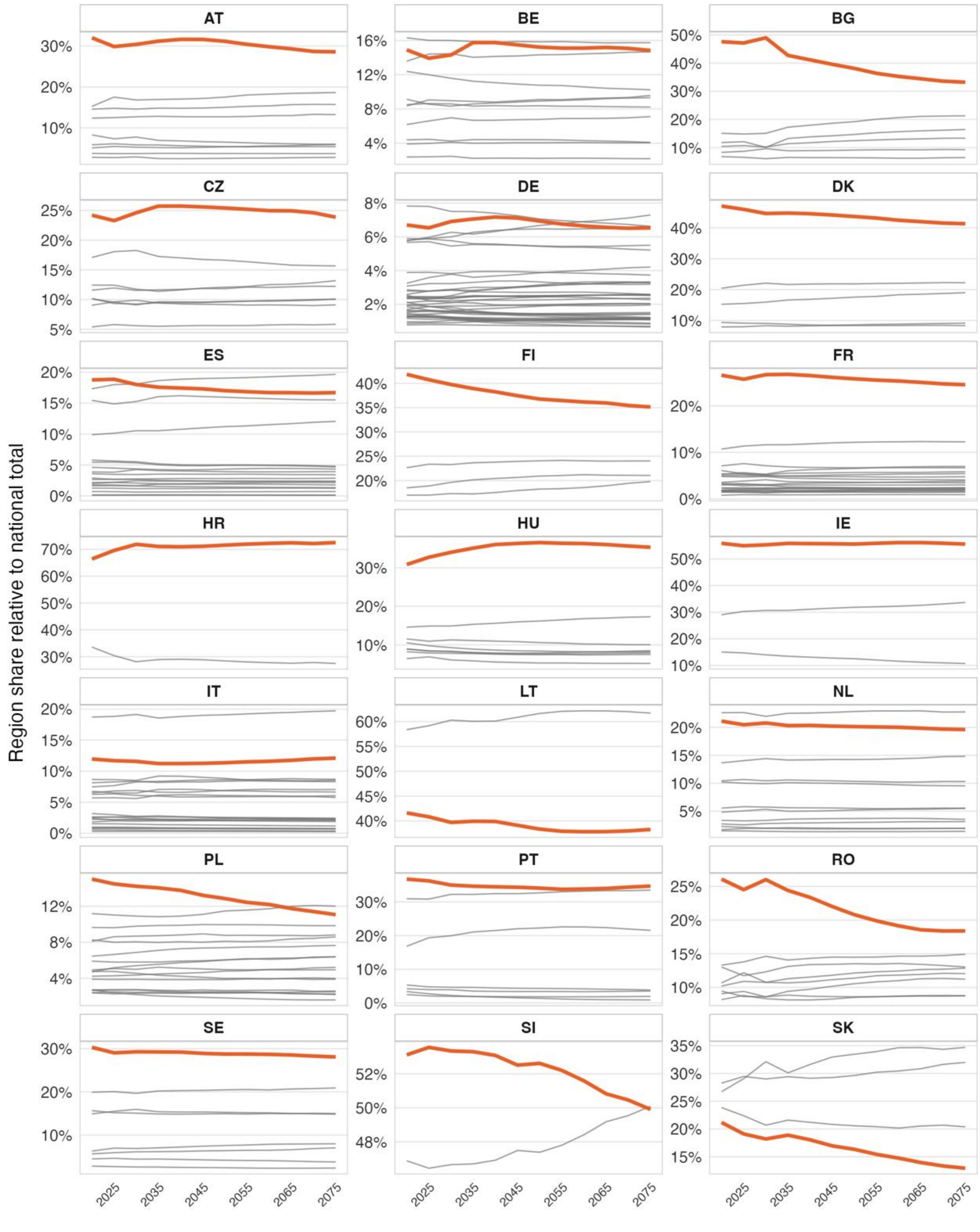

Figure A1: Share of the total highly educated population (ISCED 5-8) aged 25 to 44 held by each NUTS-2 region relative to the national total, under SSP 2, from 2015 to 2075. Each grey line represents a single NUTS-2 region; the orange line highlights the capital region. Note that y-axis scales vary across country panels to improve readability.