social sciences

MDPI

Article

Predicting Long-Run Income Growth from Women's Indicators in Developing Countries

Manuel Chaves-Maza

Department of Economy, Quantitative Methods and Economic History, Universidad Pablo de Olavide, Carretera de Utrera km 1, ES-41013 Seville, Spain; mchaves@upo.es

Abstract

Cross-national studies usually reduce women's economic position to one or two composite indices, so it remains unclear which dimensions of gender inequality are most closely associated with long-run growth in low-income economies. Drawing on 128 sex-disaggregated indicators from the World Bank Gender Data Portal for 62 low- and lower-middle-income countries over 1960–2023, we combine principal component analysis and clustering with machine-learning models to identify the health, educational, and demographic correlates of the long-run trend in gross domestic product (GDP) per capita. Under repeated cross-validation, a random forest explains 38 percent of out-of-sample variation. Health indicators occupy six of the ten most influential predictors: the share of deaths among women aged 15–59 attributable to injuries dominates under two importance measures, followed by the injury share of all female deaths, deaths from communicable and maternal causes, and female primary-school enrollment. Female schooling and child survival improved fastest where they started lowest, and those countries recorded flatter GDP per capita trends. We interpret these associations, which are predictive rather than causal, through human-capital theory, and identify preventable female mortality, female schooling, and fertility-related constraints as priorities for causal evaluation rather than proven levers of growth.

Keywords: gender inequality; female mortality; female schooling; human capital; machine learning; random forest; cross-validation

1. Introduction

The debate on women's participation in the economy has gained worldwide importance in recent years. According to many studies and international agencies, the promotion of gender equality is not only a moral necessity but also a driver of sustainable economic development (Duflo 2012; Raman et al. 2022). Capable of promoting inclusive development and thereby improving the welfare of their local communities, women entrepreneurs are acknowledged as major players of social change and innovation (Calás et al. 2009; Haugh and Talwar 2016; Setini et al. 2020; Veckalne and Tambovceva 2023). Because of ingrained gender inequality and poverty conditions, women still face greater obstacles when starting businesses, obtaining financing, and accessing adequate training in many countries (Dormekpor 2015; Hussain et al. 2019; Nisak and Sugiharti 2020).

Even though most experts agree on the broad direction of this relationship, the real-world effect of gender-related constraints on aggregate economic performance remains contested. One reason is that 'women in the economy' is typically proxied by only one or two aggregate numbers at a time. In low- and lower-middle-income countries, health deficits, restricted schooling, family pressures, and the absence of care networks arrive

Check for updates

Academic Editor: Antonella Rocca

Received: 27 September 2026
Revised: 20 October 2026
Accepted: 22 October 2026
Published: 29 October 2026

Copyright: © 2026 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

together, yet cross-national studies tend to examine each in isolation. That fragmentation makes it difficult to see how accumulated disadvantage maps onto future economic trajectories. Governments and development agencies are left with a simple but demanding question: which dimensions of social life—access to health care, education, employment, or the wider quality of institutions—are most closely associated with lasting economic change, given that they are strongly interconnected (Becker 1964; Raman et al. 2022)?

This research examines a wide set of gender-related indicators alongside growth in gross domestic product (GDP) in low-income economies between 1960 and 2023. Data come mainly from the World Bank Gender Data Portal, and the World Economic Forum Global Gender Gap Index is used only to situate the sample. A harmonized set of country-level series was built from these sources, covering health, schooling, demographic structure, labor-market conditions, and the business environment. The analytical strategy combines dimension reduction (principal component analysis) with exploratory grouping (cluster analysis) to characterize the covariance among indicators. Rather than relying on a single specification, an ensemble of learning algorithms is used to assess how well sets of indicators predict long-run expansion, and repeated k-fold cross-validation, with all predictor screening carried out inside the training folds, is used so that noise is not mistaken for signal (Breiman 2001; Friedman 2001; Chen and Guestrin 2016).

Despite a broad consensus that gender equality and economic development reinforce one another (Duflo 2012; Santos Silva and Klasen 2021), three gaps remain. First, most cross-national studies proxy 'women in the economy' with one or two aggregate indicators, typically a composite gender-gap index or female labor-force participation, which obscures which specific dimensions of women's lives—health, schooling, fertility, or demographic structure—are most closely tied to long-run growth. Second, the dominant empirical toolkit, linear regression on a handful of regressors, is poorly suited to the dense, collinear, and nonlinear web of gender indicators that low-income countries generate. Third, the evidence is concentrated in high-income or single-country settings, leaving the Global South, where the human-capital stakes are highest, comparatively under-studied (Raman et al. 2022).

This article addresses these gaps by combining a wide panel of sex-disaggregated indicators with machine-learning methods explicitly designed for high-dimensional, nonlinear prediction. Rather than testing a single coefficient, we let an ensemble of learners rank which gender-related demographic, educational, and health factors best predict the long-run trend in GDP per capita across 62 low- and lower-middle-income countries between 1960 and 2023, and we discipline that data-driven exercise with human-capital theory (Becker 1964). The objective is comparative and predictive: to rank the predictive importance of different gender-related indicators, not to identify the mechanisms through which women's health and education raise output, which would require designs able to establish temporal precedence and to address confounding (Jayachandran 2015; Santos Silva and Klasen 2021). The contribution is therefore twofold. Substantively, we identify a small set of female-specific indicators—causes of death among women of working age, schooling, and fertility—that dominate the prediction of the income trend. Methodologically, we show how interpretable ensemble learning, evaluated out of sample, can complement, rather than replace, theory-led development economics.

The remainder of the paper is organized as follows: Section 2 develops the theoretical framework linking poverty, human capital, and women's economic role. Section 3 describes the data sources, the construction of the 62-country sample, and the analytical strategy. Section 4 reports descriptive trends, the principal components and clusters, and the variable-importance ranking of the best-performing model. Section 5 interprets these findings through human-capital and endogenous-growth theory and sets out the limits of a predictive design. Section 6 concludes.

2. Theoretical Framework

2.1. Poverty, the Economy, and the Role of Women

Poverty and growth in gross domestic product (GDP) stand in a complicated and reciprocal relationship. One could argue that, by creating independent sources of income and employment, entrepreneurship is a means to reduce poverty. Conversely, conditions of extreme poverty may make it difficult for businesses to grow and survive. Many studies have asked whether GDP per capita in fact helps to reduce income inequality and poverty. At the macro level, there is evidence that nations with more entrepreneurial activity often show lower degrees of economic disparity (Bonaparte and Kumar 2024). Quantitatively, appreciable declines in income inequality have been linked to a notable rise in entrepreneurship rates, which implies that support for new businesses, especially those with growth potential, may constitute a successful approach to more inclusive development (Bonaparte and Kumar 2024). Not all kinds of business formation, however, contribute equally to lowering poverty. While necessity self-employment, which is common among poor populations without job alternatives, exerts only a limited effect on poverty reduction, research in transition economies indicates that firms with high growth and job-creation potential tend to improve individual economic well-being (Korosteleva and Stępień-Baig 2020).

That is to say, it is unlikely for someone to escape poverty if they launch a business simply out of a lack of alternative options, since such ventures are often low-productivity and yield marginal incomes. On the other hand, when the national economy evolves favorably and growth follows from a recognized opportunity accompanied by specific capital, skills, and market access, the chances of raising living standards are greater (Korosteleva and Stępień-Baig 2020).

An increase in the number and dispersion of small businesses without any significant improvement in poverty indicators could result either from the informality of the new businesses or from their being merely subsistence-based. Poverty also shapes access to the basic resources these new businesses need. Low-income individuals often find it difficult to obtain technical training, financing, productive inputs, and even the most basic business requirements: electricity, internet, and transportation. In this respect, the lack of access to conventional credit has driven the development of microcredit programs aimed at the poor, and especially at poor women, with the dual objective of promoting productive activities and economically empowering them. Although the impact on other dimensions of human poverty, such as children's education and family health, is less immediate, the evidence on microfinance is encouraging but complex: in Pakistan, for example, women who accessed larger microcredit loans were found to reduce their financial poverty, measured in income and assets, more substantially (Hussain et al. 2019).

This implies that women involved in the economy who have access to capital through microloans can invest in their small businesses and raise their income, and can therefore progressively escape financial poverty. To show notable changes in other domains, however, elements including health and educational well-being may need more time or additional investment. Microfinance experience around the world nevertheless converges on the finding that credit to low-income women, combined with training and support, can empower these women and help them move their businesses beyond a merely subsistence level toward a more sustainable one (Hussain et al. 2019). Empirical research further shows that entrepreneurship among low-income individuals is severely limited by a lack of access to credit and training programs, which feeds poverty cycles (Ogujiuba et al. 2021; Malisetti and Singh 2024).

Ogbari et al. (2024) show how empowering women entrepreneurs has far-reaching effects in settings such as Nigeria: women-led businesses not only raise household incomes

and reduce poverty, but also challenge gender stereotypes by demonstrating women's capabilities in the business world.

Public services and infrastructure constitute another vital resource. Particularly in rural or underdeveloped areas, poor access to services including electricity, drinking water, transportation, or the internet often aggravates poverty conditions and limits entrepreneurial possibilities. Case studies show that granting access to energy can trigger multiplier effects: the availability of dependable electricity enables higher productivity of small businesses, extended operating hours, and even the diversification of economic activities (Pueyo and Maestre 2019). It has also been noted, however, that women may gain less from these developments if gender disparities are not addressed. Because women typically run smaller-scale, less energy-intensive businesses, such as home workshops and local trade, general electrification programs may overlook the particular needs of their enterprises (Pueyo and Maestre 2019). These entrepreneurship and microfinance studies describe individual- and firm-level processes (Elitzur and Solodoha 2021; Solodoha et al. 2026); they are reviewed here as context for women's economic participation and are not represented as a mechanism in the country-level models of Section 3.

2.2. Health, Demographic, and Educational Factors

Apart from the socio-economic and policy elements discussed above, demographic, educational, and health conditions shape the dynamics of economic development. These elements help define the human setting in which new economic activity develops, and their influence can be direct or indirect. From a demographic standpoint, the age distribution and sex ratio of the population can influence entrepreneurial activity. An older population with accumulated experience can support more sophisticated, consulting-based ventures, whereas a large young population is usually associated with higher necessity-driven entrepreneurship when the economy fails to generate a sufficient number of formal jobs.

Our dataset includes indicators of the female population in specific age groups, such as 0 to 14 years and 55 to 59 years of age. While a high share of young people suggests many potential future entrepreneurs, it can also point to present difficulties in education and employment; conversely, a sizable group approaching retirement could indicate the availability of seasoned mentors, or of individuals who choose to launch a new venture after concluding a first career.

According to the literature, settings with an abundance of young people require major investments in education and training to channel demographic potential toward productive activity rather than informal occupation (Raman et al. 2022). Women's entrepreneurial activity, by contrast, tends to be lower where high fertility from a very young age combines with heavy household responsibilities. Reflecting low-quality entrepreneurship linked to a lack of opportunities, a study in transition economies found that women lacking job alternatives may be driven into self-employment as a last resort (Korosteleva and Stępień-Baig 2020).

This result is a reminder that demographic factors, such as family size and average age at motherhood, interact with business development: very large families, for example, can limit the time and resources a woman devotes to a business, although in some cases they also provide family labor that supports microenterprises. Regarding human capital and education, it is well established that schooling improves entrepreneurial capability. Higher levels of education are linked to cognitive, technical, and managerial abilities that help individuals recognize business prospects, formulate plans, and negotiate market obstacles. Indicators such as the enrollment rate in pre-primary education and expected years of schooling are included in our analysis, reflecting the role of an educational foundation running from early childhood to youth. Better later academic performance and, over time,

a more skilled and creative workforce are the expected returns to quality pre-primary education.

Likewise, measures such as learning-adjusted years of schooling, which combine the quantity and quality of education received, reflect the actual level of competencies in the population; a high value implies stronger human capital, a vital factor for the emergence of competitive enterprises and the attraction of investment. Gender equity in education deserves particular attention: when girls receive the same education as boys, the pool of talent available for future entrepreneurship is doubled. Achieving parity in secondary and higher education has contributed to more women acquiring business and technological skills, thereby increasing their presence in start-up ecosystems, especially in developed economies (Raman et al. 2022). Higher education and entrepreneurial training are consistently linked to greater business success; Aparicio et al. (2022) and Ilie et al. (2021), for instance, report that women with higher levels of education are more confident and more skilled in starting sustainable businesses.

Moreover, the presence of women in prominent educational roles, such as primary and higher-education teachers, can positively influence the aspirations of girls and the construction of more egalitarian societies that value women's role in the economy. It is no coincidence that one of the indicators analyzed is the proportion of female primary-school teachers: this factor can be read as a proxy for female participation in the education sector, which often correlates with favorable attitudes toward women's capabilities and can translate into their economic and entrepreneurial participation. In sum, quality education for the entire population and the removal of gender gaps in education generate the human capital necessary for a prosperous economy.

Family health is a further influential factor: when children and other dependents enjoy good health and access to care, women can devote more attention to productive activity instead of spending most of their time caring for the sick. Studies indicate that microcredit programs, by empowering women and improving their incomes, end up positively affecting family health and children's education, and hence create a virtuous circle of well-being (Hussain et al. 2019). Improvements in well-being, whether through public policy or through income per capita, feed back into economic performance by freeing people from basic survival concerns. In sum, demographic, educational, and health factors shape the backdrop against which economic development unfolds.

2.3. Human Capital and the Gendered Foundations of Growth

Taken together, the strands above converge on a single theoretical anchor: human-capital theory (Becker 1964). In its canonical form, the theory holds that investments in the health and education of a population raise productivity and, through that channel, per capita output. Its gendered extension is straightforward but consequential. Because women bear and raise children, transmit health and cognitive endowments to the next generation, and constitute roughly half of the potential labor force, under-investment in women's human capital depresses growth twice over: once through the foregone productivity of women themselves, and again through the diminished human capital of the children they raise. Endogenous-growth models reinforce this logic by treating the stock of human capital, rather than physical capital alone, as the engine of sustained increases in income. The theoretical literature identifies four channels through which gender gaps in health and education affect growth: human-capital accumulation, labor productivity, demographic change through lower fertility, and women's participation in market activity (Duflo 2012; Santos Silva and Klasen 2021). Their strength depends on the institutions and social norms that shape women's opportunities, which differ widely across developing countries (Jayachandran 2015).

This framing yields three expectations that structure our empirical work. First, indicators of women's health, particularly preventable and injury-related mortality and maternal risk, should be among the strongest correlates of long-run growth, because ill health removes workers from production and diverts household and public resources toward care. Second, female schooling should be positively associated with growth, both directly, through a more skilled workforce, and indirectly, through lower fertility and healthier children. Third, demographic variables that capture the pace of the fertility transition should matter, because a falling youth-dependency ratio frees resources for productive investment, the so-called demographic dividend. These expectations do not presuppose any particular functional form; they simply indicate which constructs a credible model should surface as important. Because the empirical work is predictive, the expectations are checked against the ranking of predictors; the channels themselves are not tested. We therefore turn to methods able to recover such relationships from a large, collinear indicator space without imposing linearity in advance.

3. Materials and Methods

3.1. Information Sources

To assemble information on indicators of gender equality and economic development, this study combines data from several reliable sources. First, a wide range of sex-disaggregated development indicators is collected from the World Bank Gender Data Portal (World Bank 2023). These include variables on female educational attainment (for example, expected years of schooling for girls), health outcomes (for example, maternal mortality rates and female life expectancy at birth), economic status (for example, female unemployment and labor-force participation), and representation (for example, women's share of ministerial positions and parliamentary seats). The analytical dataset contains 128 indicators grouped into health (39), education (22), labor and business (26), population and demography (33), and empowerment-related (8) domains; Supplementary Table S1 lists each indicator with its World Bank code and temporal coverage.

Second, information from the World Economic Forum Global Gender Gap Report 2023 is used to incorporate a comprehensive assessment of gender inequality (World Economic Forum 2023). The Global Gender Gap Index (GGGI) is a widely used composite index that evaluates gender parity annually in four domains: political empowerment, health and survival, educational attainment, and economic participation and opportunity (World Economic Forum 2023). The index assigns each country a score ranging from 0 to 1, where 1 denotes complete parity, together with a rank among the countries assessed. Covering 146 countries, the 2023 report offers a current picture of national progress toward closing the gender gap. The GGGI is used only as a contextual reference: it places the sample in a global context but does not enter any model, and it is not used to validate the findings, because its sub-indices overlap conceptually with the World Bank predictors.

3.2. Sample, Data Processing, and Cleaning

To prepare the data for analysis, countries were filtered according to the availability of the key indicators over the period of interest (1960–2023), retaining only low-income and lower-middle-income economies under the World Bank income classification. This yielded 62 countries with a complete set of trend estimates for all 128 indicators, so the final dataset contains no missing values; the countries are listed in Supplementary Table S2.

Harmonization of indicators. A number of indicators had to be harmonized with respect to time. Each indicator was used in the units published by the World Bank. We further ensured that higher values always translated into 'more' of the relevant variable, though not always into a better outcome; a higher unemployment rate, for instance, was

a worse result. This check was essential before estimating trends, in order to prevent misinterpretation arising from scale inversion or from evolving definitions.

Analysis of trends (1960–2023). For each country and indicator, the long-run trend is the ordinary least squares slope b in the regression $x(t) = a + b \cdot t + e(t)$, estimated on all years t between 1960 and 2023 for which the World Bank reports a value. The slope is expressed in units of the indicator per year: percentage points per year for rates and shares, births per woman per year for fertility, and persons per year for population and labor-force counts. The outcome is the slope of GDP per capita in current US dollars (World Bank code in Supplementary Table S1), that is, the average absolute change in GDP per capita, in US dollars per year, over the years observed; the sample mean of 30.6 means that GDP per capita rose by about 31 current US dollars per year in the average country. The outcome is therefore an estimated linear trend, not a percentage growth rate or a difference between endpoint values. Because the series are in current dollars, the trend also reflects price and exchange-rate movements, and because it is expressed in absolute terms it is larger, other things equal, for countries with higher income levels. Coverage differs across indicators: GDP per capita is observed for a median of 64 years per country, whereas the cause-of-death indicators are modelled estimates available for six years between 2000 and 2021 (Supplementary Table S1); for those indicators the slope describes change since 2000.

Missing-data management. No values were imputed. Slopes were estimated from the annual observations available for each country and indicator, so gaps at the start, middle, or end of a series reduce the number of observations behind a slope rather than being filled. Indicators too incomplete to support a slope for every country were omitted rather than imputed, which is why the final dataset contains no missing values. The number of annual observations per indicator is reported in Supplementary Table S1.

Normalization and outlier checks. Prior to analysis, the distributions of the indicators and of their trends were reviewed for outliers or skewed values that might unduly influence results. In a few cases extreme outliers were noted, for example, countries with unusually large swings in certain rates, such as a very high increase in unemployment or a sharp drop in business rank, arising from conflict or economic crisis. No observation was removed; instead, Section 4.3 reports a sensitivity analysis that excludes the three countries with the most extreme outcome values (Angola, Samoa, and Guinea). All predictors were standardized to zero mean and unit variance before principal component analysis, clustering, and the scale-sensitive learners (support vector machine, neural network, and ridge regression); tree-based models used unscaled values.

3.3. Analytical Techniques

To examine and summarize the data, the study uses several complementary analytical approaches: principal component analysis, cluster analysis, and supervised machine learning, chiefly random forests. As explained below, each technique provides a distinct route to identifying patterns and relationships in the dataset.

Cluster analysis is an unsupervised classification technique that groups related observations according to their attributes, without predetermined labels (Frades and Matthiesen 2010; Song et al. 2010). In exploratory contexts, such as geophysical data interpretation where underlying categories are unknown beforehand, this method identifies natural groupings in the data (Song et al. 2010).

Factor analysis condenses the information in numerous variables into a few composite dimensions that account for a substantial share of the variance, either by extracting a smaller number of factors or by applying principal component analysis (PCA), a related technique (McVean 2009). Factor loadings, which show how strongly each original variable is associated with each factor, are used to interpret the resulting factors or components

(Izquierdo et al. 2014). To guarantee that the solution is reliable and reproducible, factor analysis must be applied carefully, which calls for meticulous reporting and deliberate choices regarding the extraction technique, the number of factors retained, and rotation (Izquierdo et al. 2014). Researchers must consider the validity and reliability of derived factors, since failing to do so may result in the interpretation of factors that are statistical artifacts or that do not generalize (Armstrong and Soelberg 1968; Izquierdo et al. 2014).

Principal component analysis was applied to the standardized trends of the 96 rate, ratio, and index indicators. The 32 indicators measured as absolute counts (population by age group, the female labor force, and numbers of deaths or of children out of school) were excluded because their trends mainly reflected country size, and the outcome was not included. The first ten components were retained as inputs to k-means clustering with 50 random starts. The number of clusters was chosen by comparing the within-cluster sum of squares (elbow) and the average silhouette coefficient for k = 2 to 6 (Rousseeuw 1987), and stability was assessed by re-estimating the solution on 200 bootstrap samples. As a sensitivity check, the analysis was repeated with all 128 indicators, with and without the outcome. This step is descriptive: components and clusters summarize how the indicators co-vary and are not used as inputs to the predictive models.

In this analysis, several machine-learning models were used to predict the GDP per capita trend on the basis of a wide set of 128 gender-related indicators spanning health, education, labor, and demographic factors. The random forest, an ensemble tree-based method, was the primary candidate; we also compared gradient boosting, XGBoost, a support vector machine (SVM), a feed-forward neural network, and a linear ridge regression as a benchmark. Random forests (Breiman 2001) use bagging to train many decision trees on bootstrapped samples and introduce randomness into feature selection in order to improve generalization. Ensemble methods of this kind are known to reduce overfitting by averaging over several learners, which makes them well suited to detecting nonlinear relationships. Gradient boosting (Friedman 2001) and its refinement XGBoost (Chen and Guestrin 2016) construct boosted ensembles by training trees sequentially so that each corrects the errors of its predecessors. SVMs (Cortes and Vapnik 1995), by contrast, seek an optimal separating hyperplane in a high-dimensional space for regression and classification, while deep neural networks learn multi-layer feature representations from data (LeCun et al. 2015). We adopted an ensemble focus because theory and prior research show that combining predictors can make predictions more accurate and more stable; bagging and boosting frequently outperform single models by exploiting diverse learners (Breiman 2001; Friedman 2001). The validation procedure is described in Section 3.4.

3.4. Validation and Interpretation

Interpretation focuses on both the static relationships among variables and their dynamic behavior over time, with particular attention to each variable's slope or trend. The slope is the coefficient obtained by regressing the variable on time, and it measures the rate of change per unit of time (Roustaei 2024; Sykes 1993).

The slope of each variable over time is a further feature that can be used in cluster analysis to describe the observations being clustered. Consequently, cases that display comparable trend directions or magnitudes throughout the study period may be grouped together (Song et al. 2010; Frades and Matthiesen 2010). Because cluster analysis is exploratory, external validation, for example, mapping clusters to known categories and applying domain knowledge, is frequently required to verify that trend-based clusters make theoretical and practical sense (Frades and Matthiesen 2010; Song et al. 2010).

Because our aim is to compare the predictive importance of indicators rather than to estimate a single structural coefficient, model performance was assessed on countries the

model had not seen. Each algorithm was evaluated with 5-fold cross-validation repeated 10 times (50 train–test splits). To avoid selection bias (Ambroise and McLachlan 2002), the screening step that retains the 15 indicators most correlated with the outcome was carried out inside each training fold, never on the full sample, and scaling was likewise fitted on training data only. Hyperparameters were fixed in advance (random forest: 500 trees; gradient boosting and XGBoost: 200 trees, learning rate 0.05, maximum depth 3; SVM: radial kernel, C = 10; neural network: one hidden layer of 50 units, L2 penalty 0.1; and ridge regression: penalty chosen by internal cross-validation over 25 values between 0.001 and 1000). For the random forest we also report nested cross-validation, in which a grid search over the number of trees, tree depth, and minimum leaf size is run inside each training fold of a 5-fold cross-validation repeated 5 times (Varma and Simon 2006). Performance is summarized by the out-of-fold R^2, computed by pooling the predictions for all 62 countries within each repetition, and by its range across the ten repetitions; leave-one-out cross-validation is reported as a check. Variable importance was measured in two ways: the impurity-based importance of the random forest fitted to all 62 countries, and permutation importance on the held-out fold of each split, defined as the mean fall in R^2 when the values of one indicator are randomly shuffled (20 permutations per indicator and split) (Molnar 2020). Because impurity-based importance can favor some predictors and splits credit among correlated ones (Strobl et al. 2007), only rankings that agree across both measures are interpreted. All analyses were run in Python 3.13 with scikit-learn 1.9 and XGBoost 3.4; figures were drawn with Matplotlib 3.10. We treat the resulting importances as evidence of predictive association, not of causal effect; Section 5 returns explicitly to this distinction and to the threats—omitted variables, reverse causality, and measurement error in low-capacity statistical systems—that bound any cross-national inference of this kind.

Use of generative artificial intelligence. Generative AI tools were used for English-language editing and to support the coding of the robustness analyses; the author checked all outputs and takes full responsibility for the content.

4. Results

4.1. Descriptive Statistics and Trends, 1960–2023

Since the slope of GDP per capita measures the average absolute annual change in GDP per capita over the study period, we begin the descriptive analysis by examining how this variable is distributed across the 62 countries. The predictors fall into five broad groups: demographic, health, education, labor and business, and empowerment-related. Table 1 summarizes the descriptive statistics for the main outcome and the core gender-related predictors, together with the basic business-environment and demographic variables that characterize the sample.

The distribution of the outcome is positively skewed (skewness 0.95; mean 30.6 and median 21.1 US dollars per year). Most countries are concentrated between 0 and 20 (27 countries) and between 20 and 40 (17 countries); fewer record steeper trends, and two exceed 80 (Samoa and Angola). Only Guinea has a negative trend.

Correlation analysis identified the strongest gender–economy associations (Figure 1). Of the 128 indicators, 26 are correlated with the outcome at the 5 percent level (Pearson), against about six expected by chance, before any correction for multiple testing. All eleven significant education indicators are negatively correlated with the GDP per capita trend: countries in which female enrollment and gender parity in schooling rose fastest recorded flatter income trends. This pattern is consistent with a base effect. Female gross primary enrollment in the first year observed (median 1971; World Bank data, Code S1) is strongly negatively correlated with its subsequent trend (Spearman $\rho = -0.83$) and positively correlated with the GDP per capita trend ($\rho = 0.52$); likewise, female under-five mortality

fell fastest where it started highest (ρ = −0.84). The strongest bivariate correlation is that of female gross primary enrollment (r = −0.489; Spearman ρ = −0.558), followed by the shares of deaths among women aged 15–59 due to communicable, maternal, and nutritional causes (r = 0.470) and due to injuries (r = −0.460). Pearson and Spearman coefficients agree in sign for all fifteen leading indicators.

Table 1. Descriptive statistics of the outcome variable and gender-related indicators (n = 62 low- and lower-middle-income countries; country-level trends per year).

Category	Variable	Definition	Mean	Std. Dev.
Outcome	GDPpc_Trend	GDP per capita, current US$ (US$ per year; 1960–2023)	30.627	24.857
Business environment	Time_Start_Business	Time required to start a business (days per year; 2003–2019)	−2.636	4.016
	StartUp_Cost_F	Cost of business start-up procedures, female (% of gross national income per capita, per year; 2003–2019)	−12.036	18.948
Women's health	Injury_Deaths_F_15_59	Share of deaths among women aged 15–59 due to injuries (percentage points per year; 2000–2021)	0.070	0.224
	Maternal_Mortality	Maternal deaths per 100,000 live births (per year; 1985–2023)	−12.993	12.429
	LifeExp_F_Birth	Female life expectancy at birth (years per year; 1960–2023)	0.371	0.161
Women's education	Exp_Years_School_F	Expected years of schooling, female (years per year; 2010–2020)	0.153	0.091
	Sec_Enroll_F_Gross	Female gross secondary-school enrollment (percentage points per year; 1971–2023)	0.940	0.491
	Preprim_Enroll_F_Gross	Female gross pre-primary enrollment (percentage points per year; 1983–2020)	1.555	5.280
Education system	Primary_Teachers_F_pct	Female share of primary-school teachers (percentage points per year; 1971–2023)	0.448	0.355
Labor market	LFP_F_15plus	Female labor-force participation, ages 15+ (percentage points per year; 1990–2023)	−0.127	0.326
	Unemp_F	Female unemployment rate (percentage points per year; 1991–2023)	0.033	0.211
	LaborForce_F	Female labor force (persons per year; 1990–2023)	87,559	137,827
Demography	Fertility_Total	Total fertility rate (births per woman per year; 1960–2023)	−0.055	0.028
	Adolescent_Fertility	Births per 1000 women aged 15–19 (per year; 1960–2023)	−1.120	0.712
	Pop_F_0_14	Female population aged 0–14 (persons per year; 1960–2023)	76,245	118,221
	Pop_F_55_59	Female population aged 55–59 (persons per year; 1960–2023)	6109	9712
	Pop_F_Total	Total female population (persons per year; 1960–2023)	212,658	314,094
Political empowerment	Women_Parliament_Seats	Women's share of parliamentary seats (percentage points per year; 1997–2023)	0.655	0.533
Demography	Age_First_Marriage_F	Women's age at first marriage (years per year; 1975–2015)	0.059	0.050

Notes: All statistics summarize the country-level ordinary least squares trend (slope per year) of each indicator over the years observed between 1960 and 2023; count variables (female labor force and population) are trends in persons per year. GDP per capita is in current US dollars. The years covered by each indicator are given in parentheses (median first and last year across countries; Supplementary Table S1). Source: World Bank Gender Data Portal and World Development Indicators.

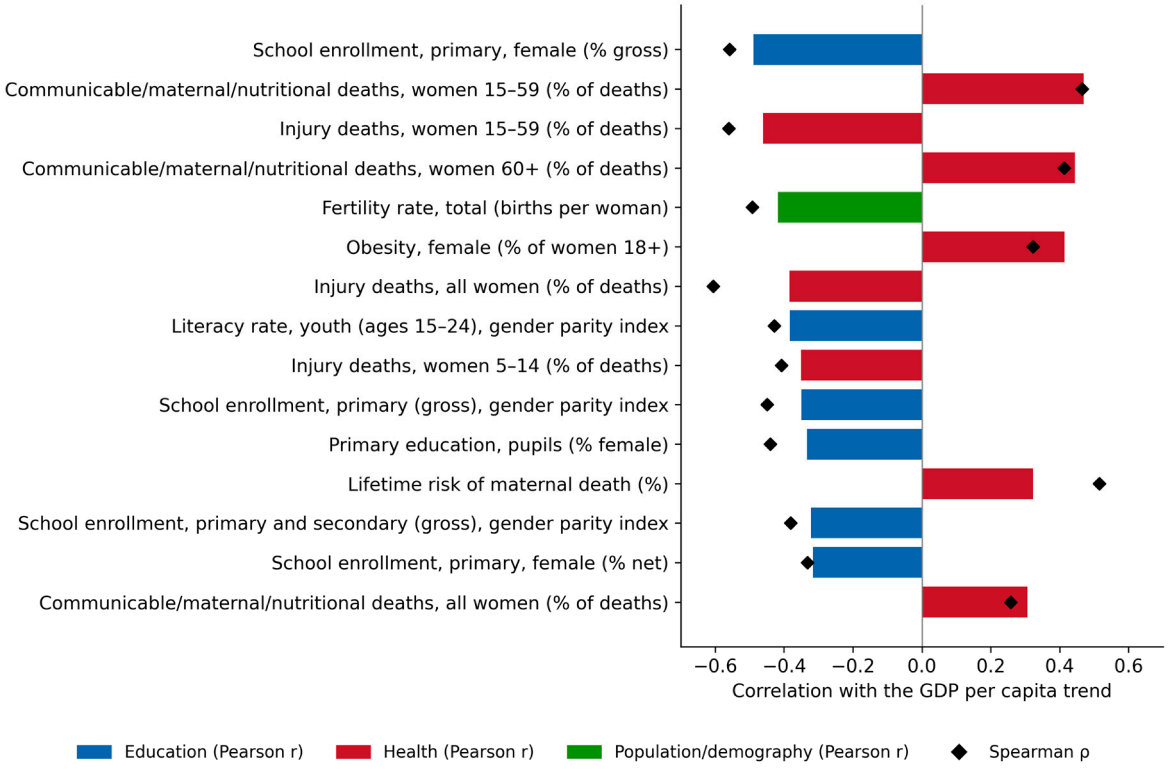


Figure 1. The fifteen indicators most strongly correlated with the GDP per capita trend (bars, Pearson correlation coefficient r; diamonds, Spearman rank correlation coefficient ρ) in 62 countries.

4.2. PCA and Cluster Analysis

In addition to examining individual factors, the analysis investigated whether countries can be grouped into distinct categories according to their gender-related development profiles. Principal component analysis of the 96 rate, ratio, and index indicators (Section 3.3) shows that no single component dominates: the first two components explain 15.6 and 9.7 percent of the variance and the first ten together 64.5 percent, so the indicators capture several distinct dimensions rather than one underlying construct. The first component loads on the pace of expansion in female schooling and in gender parity in enrollment and literacy, together with gains in female life expectancy and falls in under-five mortality (loadings in Supplementary Table S3). The within-cluster sum of squares falls smoothly with k, without a clear elbow, and the average silhouette is low for every k between 2 and 6 (0.13–0.14), so the countries lie along a continuum more than in separate groups (Figure 2).

For descriptive purposes, the two-cluster solution separates 21 countries that combined slower expansion in female schooling with steeper GDP per capita trends (median 40.0 US dollars per year) from 41 countries, mostly in sub-Saharan Africa and South Asia, that recorded the fastest gains in female enrollment and child survival, typically from the lowest starting points, and flatter income trends (median 16.3; Mann–Whitney $p = 0.006$). The outcome was not used to form the clusters. The first component is negatively correlated with the outcome (Spearman $\rho = -0.31$, $p = 0.014$).

Two checks qualify this picture. The two-cluster assignment is unstable: re-estimated on bootstrap samples, it agrees only weakly with the full-sample solution (mean adjusted Rand index 0.29). When the 32 count indicators are kept, the first component (20.4 percent of the variance) reflects absolute population growth, and the two-cluster solution simply separates the six most populous countries (Bangladesh, Egypt, Ethiopia, Nigeria, Pakistan, and the Philippines) from the rest, with no difference in the outcome ($p = 0.39$); adding the outcome to the clustering variables leaves that solution unchanged (adjusted Rand index

1.00). The clusters are therefore a descriptive summary of how the indicators co-vary, not evidence of distinct development pathways.

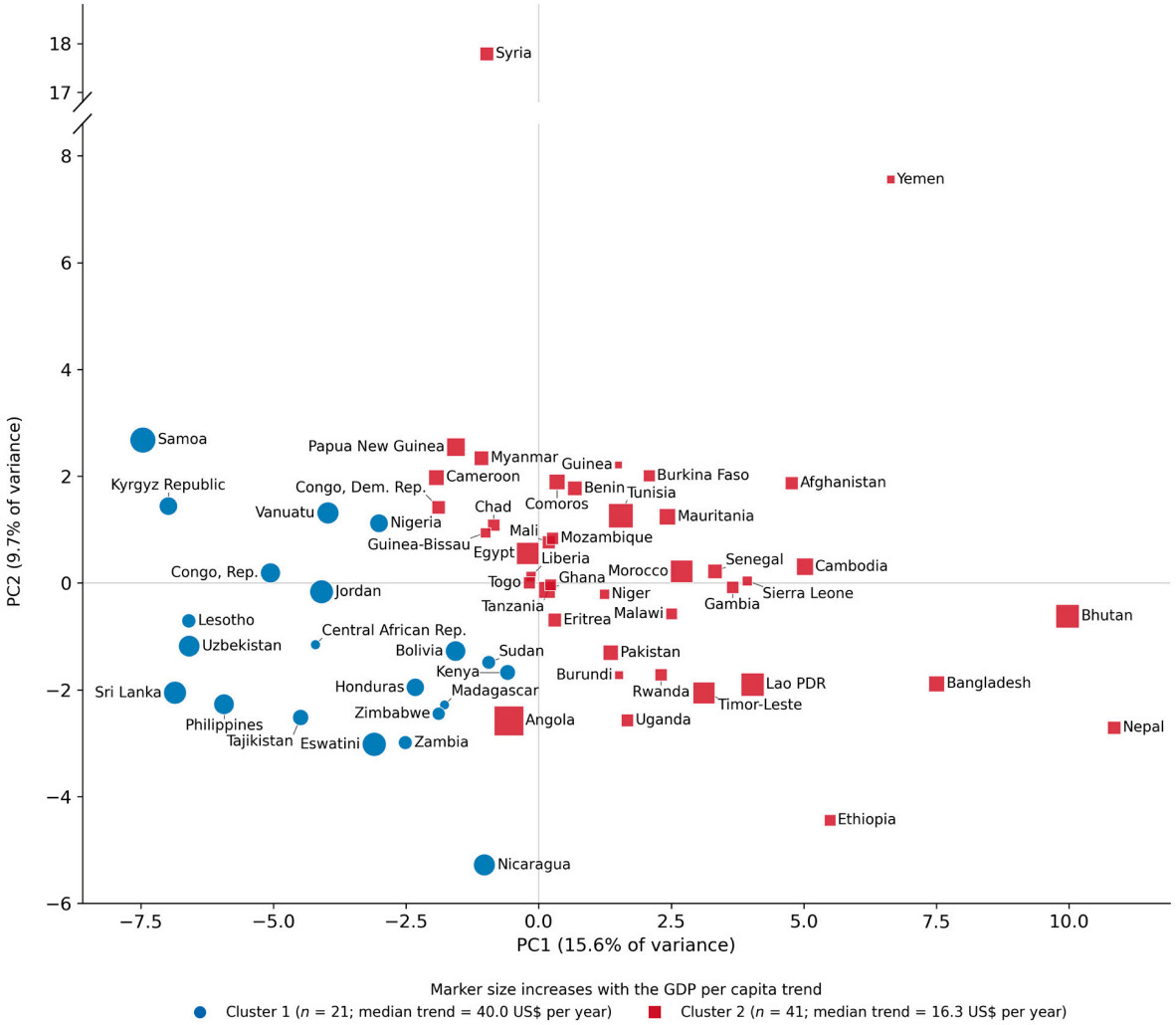


Figure 2. Countries on the first two principal components of the 96 rate, ratio, and index indicators, by k-means cluster (k = 2); marker size increases with the GDP per capita trend; the vertical axis is broken to show the Syrian Arab Republic.

4.3. Random Forest and Comparison of Machine-Learning Models

Under repeated cross-validation the random forest performed best among the models tested, with an out-of-fold R^2 of 0.38 (range across the ten repetitions 0.28–0.49; leave-one-out R^2 0.38). XGBoost (0.30), gradient boosting (0.29), the linear ridge benchmark (0.29), and the SVM (0.24) followed, while the neural network predicted worse than the sample mean ($R^2 = -0.80$). Mean fold-level R^2 for the random forest was 0.28 with fixed hyperparameters, 0.29 with nested tuning, 0.29 when all 128 indicators were used without screening, and 0.39 when the three countries with the most extreme outcomes were excluded. The gain in the random forest over the linear benchmark is therefore modest, and about 60 percent of the cross-country variation in the outcome remains unexplained. Accuracy also depends heavily on how the sample is split: across 100 random 70/30 splits, the test-set R^2 of the random forest ranged from −0.36 to 0.78 (mean 0.35), which is why we report repeated cross-validation rather than a single hold-out score.

Beyond raw performance, the random forest provides information on which features drive predictions through its variable-importance measures (Figure 3). Health-related variables predominate: health indicators account for six of the ten best predictors under

both measures. The trend in the share of deaths among women aged 15–59 attributable to injuries was the most important predictor: it passed the screening step in all 50 training folds, it accounts for 40 percent of impurity-based importance in the full-sample model, and shuffling it lowered out-of-fold R^2 by 0.43 on average. Countries in which this share rose faster recorded flatter GDP per capita trends. The next predictors, in the same order under both measures, were the injury share of all female deaths, the share of deaths among women aged 15–59 from communicable, maternal, and nutritional causes, and female gross primary enrollment. Beyond these four, the two measures diverge (Spearman rank correlation across all screened indicators 0.07) and permutation importance is close to zero, so lower-ranked indicators should not be ordered.

Female education variables also rank among the leading predictors, but their association with the outcome is negative, which runs against the second expectation in Section 2.3 when read at face value. The likeliest explanation is the base effect documented in Section 4.1: the slope measures the pace of change, enrollment rose fastest where it started lowest, and countries that started with lower enrollment recorded flatter GDP per capita trends. The dominance of health indicators is consistent with the first expectation, but the leading health indicators are shares of deaths by cause, which shift mechanically as communicable diseases recede and are derived from modelled estimates available only since 2000. The ranking therefore indicates which indicators carry predictive information about the income trend, not where targeted improvements in female health and education would carry the largest economic payoff.

Plotting out-of-fold predictions against observed values for each country allows the prediction quality of the random forest to be assessed visually (Figure 4). Predictions and observations are correlated (r = 0.65), but the model compresses the range: its predictions lie between 9 and 69 US dollars per year, against observed values between −17 and 112. The largest errors are for Angola (observed 111.7, predicted 35.0), Burundi (2.7 and 62.3), the Lao People's Democratic Republic, and Guinea, whose trends lie far from those of countries with similar indicator profiles.

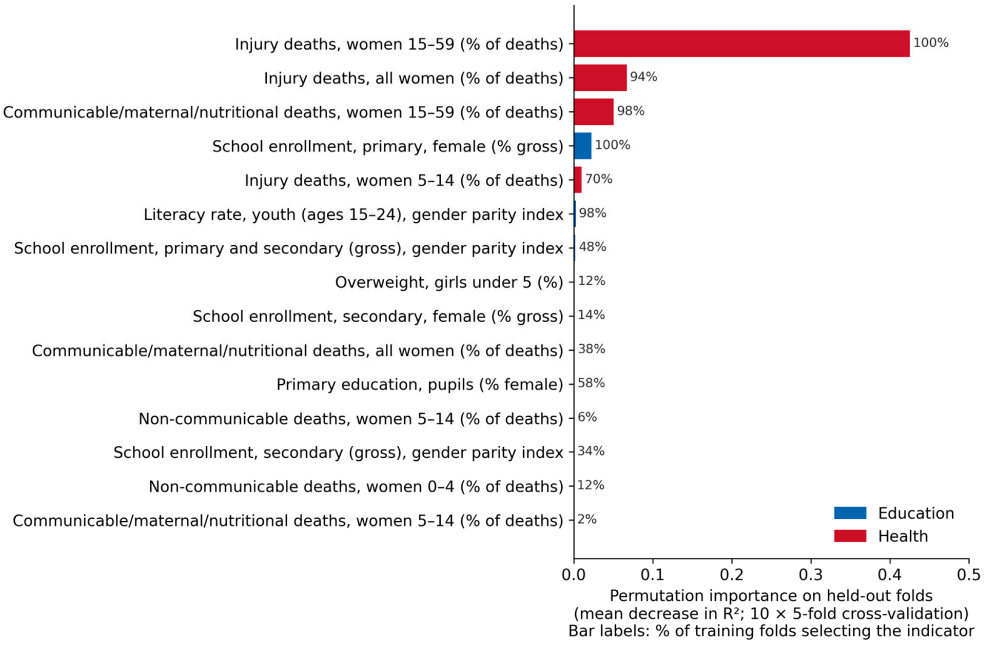


Figure 3. Permutation importance of the fifteen leading predictors on held-out folds (random forest, 10 × 5-fold cross-validation); labels give the share of training folds in which each indicator passed the screening step.

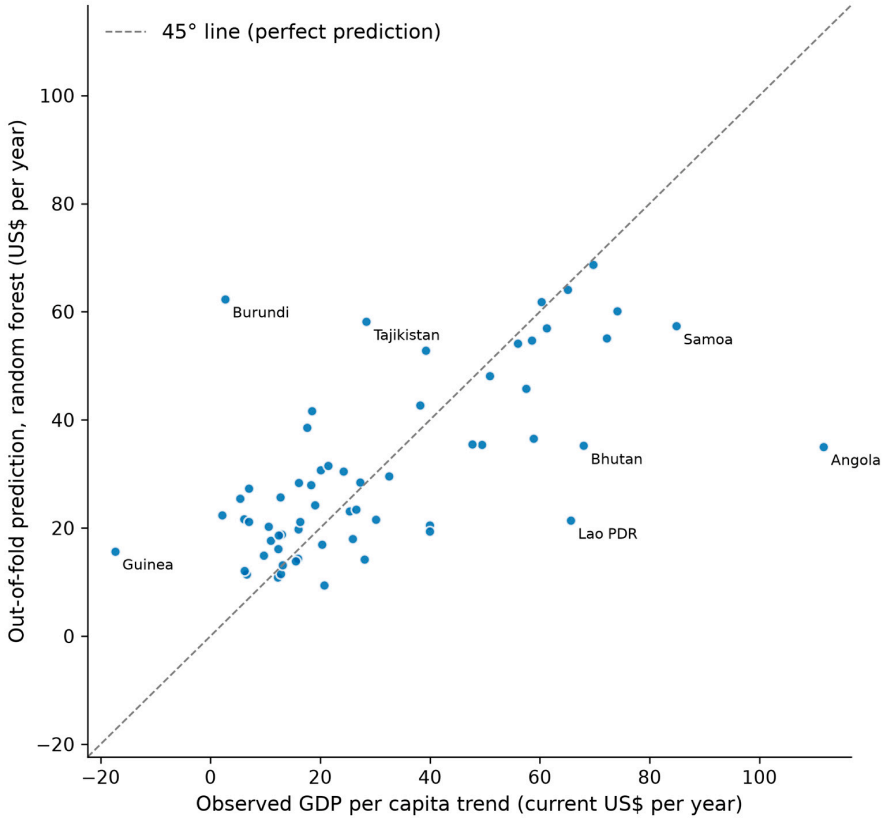


Figure 4. Random forest: out-of-fold predictions versus observed GDP per capita trends (current US dollars per year), averaged over ten repetitions of 5-fold cross-validation.

5. Discussion

This study ranked 128 sex-disaggregated indicators by their power to predict the long-run GDP per capita trend of 62 low- and lower-middle-income countries. Evaluated out of sample, a random forest explains 38 percent of the cross-country variation, only modestly more than a linear ridge benchmark (29 percent), so the advantage of nonlinear learners over linear models, one of the motivations given in the Introduction, is small in a sample of this size. Four indicators carry most of the predictive information: the injury shares of deaths among women aged 15–59 and among all women, the share of deaths among women aged 15–59 from communicable, maternal, and nutritional causes, and female primary enrollment.

The cluster analysis adds a descriptive reading to the variable-level results. Countries do not fall into sharply separated groups; they lie along a gradient in the pace of improvement in female schooling and child survival, and the countries that improved fastest, from the lowest starting points, had the flattest absolute income trends. Read together with the negative education correlations, this pattern suggests that the indicators capture where countries started as much as where they are going, which is one reason the predictive rankings cannot be translated directly into policy sequencing.

By grounding interpretation in theory, we avoid a purely data-mining or reductionist reading. Correlation is not causation, and the model by itself cannot establish why these relationships exist. The results are partly consistent with human-capital theory, since health indicators dominate the predictions, and partly at odds with it, since the education indicators enter with negative signs; theoretical consistency is therefore not, on its own, evidence that the patterns are causal.

These results can be set against cross-country studies that estimate effects rather than predictive rankings. Klasen and Lamanna (2009) find, in a panel of countries, that gender

gaps in education and employment reduce growth, and Bloom et al. (2004) estimate a positive effect of life expectancy on output in a production-function framework. Our health result is only loosely comparable with the second: our leading health indicators are cause-of-death shares, not life expectancy, and their signs do not map directly onto an effect of health on output. Our education result appears to contradict the first, but the two are not directly comparable: those studies relate growth to levels and gaps, whereas our predictors are trends whose pace depends on the starting level, as Section 4.1 shows.

The correlation between out-of-fold predictions and observed values ($r = 0.65$) shows that the random forest captures part of the signal linking the income trend and gender indicators, but most of the variation remains unexplained. The instances in which predictions deviate from observed values may reflect irreducible error or influences outside the model's feature set, such as macroeconomic shocks, natural-resource booms, or political instability not registered by gender indicators. The predictive model and an accompanying policy-recommendation tool are openly available at https://github.com/upocuantitativo/iafordevelopment (accessed on 29 October 2026; https://doi.org/10.5281/zenodo.21974290).

Several limitations bound these conclusions. First, the design is cross-sectional: predictors and outcome are trends over overlapping periods, so the associations may reflect reverse causality, with rising income financing better health and schooling, or common causes such as institutional quality, political stability, conflict, and the demographic transition, none of which are controlled for. Second, with 62 countries and 128 candidate indicators, out-of-sample accuracy is modest and sensitive to how the sample is split, and predictions for countries or periods outside the sample are untested; temporal validation is not possible because each country contributes a single trend. Third, the outcome is an absolute trend in current US dollars, which mixes real growth with price and exchange-rate movements and gives larger values to richer countries; it is not a measure of poverty, and growth in average income need not reduce poverty. Fourth, the indicators differ in coverage and quality: the cause-of-death shares that dominate the ranking are modelled estimates for 2000–2021, and measurement practices in low-capacity statistical systems changed over the period. Fifth, the relationships may differ across income groups and regions, a heterogeneity that a sample of this size cannot resolve. Replicating the analysis with constant-price outcomes, predictors measured before the outcome period, and panel designs that exploit within-country variation would show whether the ranking survives.

In practical terms, the results identify reductions in preventable female deaths, female schooling, and the fertility transition as candidates for causal evaluation; whether improving them raises growth, and at what cost, cannot be established with a predictive design.

6. Conclusions

Indicators of female human capital, particularly women's health, are among the strongest predictors of long-run income trends in low- and lower-middle-income countries. The model identified causes of death among women of working age and female primary enrollment as leading predictors of the GDP per capita trend, although it leaves most of the cross-country variation unexplained. These results partly align with human-capital theory (Becker 1964), which holds that investing in education and health raises a population's productivity and thereby economic growth; the negative sign of the education indicators shows that the alignment is not complete.

This is consistent with the broader development literature linking gender equality to growth: higher female education contributes to a more skilled labor force and to lower fertility, both of which promote increases in GDP per capita (Duflo 2012; Santos Silva and Klasen 2021). Our analysis, while data-driven, stays close to that theoretical framing but

cannot test it: the associations documented here are predictive, and the negative association between the expansion of female schooling and the income trend most likely reflects differences in starting levels rather than a negative return to education.

If the associations documented here approximate underlying mechanisms, they carry a concrete message for research and policy in low-resource settings. The prominence of preventable female deaths, both injury-related deaths among women of working age and deaths from communicable and maternal causes, suggests that programs addressing road-traffic and occupational injuries, violence against women, and maternal complications deserve rigorous evaluation of their economic as well as their health returns. Female schooling and the easing of fertility-related constraints are natural complements. These are priorities for causal research, not proven levers: their effects on growth, their costs, their feasibility, and their distributional consequences, including any effect on poverty, have to be established with designs that can identify causal effects.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/socsci15110752/s1: Table S1: Indicators, World Bank codes, and temporal coverage; Table S2: Countries, outcome values, out-of-fold predictions, and cluster membership; Table S3: Loadings of the 96 rate, ratio, and index indicators on the first three principal components; Code S1: Script reproducing the cross-validation, permutation-importance, cluster-validation, and base-effect analyses. This study draws on publicly available, country-level secondary data from the World Bank Gender Data Portal (https://genderdata.worldbank.org) and the World Economic Forum Global Gender Gap Index.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The processed dataset and the analysis code that support the findings of this study are openly available at https://github.com/upocuantitativo/iafordevelopment (accessed on 29 October 2026; https://doi.org/10.5281/zenodo.21974290). The script for the robustness analyses is provided as Supplementary Code S1.

Conflicts of Interest: The author declares no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

GDP	Gross domestic product
PCA	Principal component analysis
GGGI	Global Gender Gap Index
SVM	Support vector machine
XGBoost	Extreme gradient boosting

References

Ambroise, Christophe, and Geoffrey J. McLachlan. 2002. Selection Bias in Gene Extraction on the Basis of Microarray Gene-Expression Data. *Proceedings of the National Academy of Sciences of the United States of America* 99: 6562–66. [CrossRef] [PubMed]

Aparicio, Sebastian, David Audretsch, Maria Noguera, and David Urbano. 2022. Can Female Entrepreneurs Boost Social Mobility in Developing Countries? An Institutional Analysis. *Technological Forecasting and Social Change* 175: 121401. [CrossRef]

Armstrong, J. Scott, and Peer Soelberg. 1968. On the Interpretation of Factor Analysis. *Psychological Bulletin* 70: 361–64. [CrossRef]

Becker, Gary S. 1964. *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. New York: Columbia University Press.

Bloom, David E., David Canning, and Jaypee Sevilla. 2004. The Effect of Health on Economic Growth: A Production Function Approach. *World Development* 32: 1–13. [CrossRef]

Bonaparte, Yosef, and Naman Kumar. 2024. Does Entrepreneurship Really Reduce Income Inequality? *Finance Research Letters* 69: 106234. [CrossRef]

Breiman, Leo. 2001. Random Forests. *Machine Learning* 45: 5–32. [CrossRef]

Calás, Marta B., Linda Smircich, and Kristina A. Bourne. 2009. Extending the Boundaries: Reframing 'Entrepreneurship as Social Change' through Feminist Perspectives. *Academy of Management Review* 34: 552–69. [CrossRef]

Chen, Tianqi, and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: Association for Computing Machinery, pp. 785–94. [CrossRef]

Cortes, Corinna, and Vladimir Vapnik. 1995. Support-Vector Networks. *Machine Learning* 20: 273–97. [CrossRef]

Dormekpor, Evelyn. 2015. Poverty and Gender Inequality in Developing Countries. *Developing Country Studies* 5: 76–102.

Duflo, Esther. 2012. Women Empowerment and Economic Development. *Journal of Economic Literature* 50: 1051–79. [CrossRef]

Elitzur, Ramy, and Eliran Solodoha. 2021. Does Gender Matter? Evidence from Crowdfunding. *Journal of Business Venturing Insights* 16: e00268. [CrossRef]

Frades, Itziar, and Rune Matthiesen. 2010. Overview on Techniques in Cluster Analysis. *Methods in Molecular Biology* 593: 81–107. [CrossRef] [PubMed]

Friedman, Jerome H. 2001. Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics* 29: 1189–232. [CrossRef]

Haugh, Helen M., and Alka Talwar. 2016. Linking Social Entrepreneurship and Social Change: The Mediating Role of Empowerment. *Journal of Business Ethics* 133: 643–58. [CrossRef]

Hussain, Javed, Samia Mahmood, and Jonathan Scott. 2019. Gender, Microcredit and Poverty Alleviation in a Developing Country: The Case of Women Entrepreneurs in Pakistan. *Journal of International Development* 31: 247–70. [CrossRef]

Ilie, Camelia, Abel Monfort, Gaston Fornes, and Guillermo Cardoza. 2021. Promoting Female Entrepreneurship: The Impact of Gender Gap Beliefs and Perceptions. *SAGE Open* 11: 21582440211018468. [CrossRef]

Izquierdo, Isabel, Julio Olea, and Francisco J. Abad. 2014. Exploratory Factor Analysis in Validation Studies: Uses and Recommendations. *Psicothema* 26: 395–400. [CrossRef] [PubMed]

Jayachandran, Seema. 2015. The Roots of Gender Inequality in Developing Countries. *Annual Review of Economics* 7: 63–88. [CrossRef]

Klasen, Stephan, and Francesca Lamanna. 2009. The Impact of Gender Inequality in Education and Employment on Economic Growth: New Evidence for a Panel of Countries. *Feminist Economics* 15: 91–132. [CrossRef]

Korosteleva, Julia, and Paulina Stępień-Baig. 2020. Climbing the Poverty Ladder: The Role of Entrepreneurship and Gender in Alleviating Poverty in Transition Economies. *Entrepreneurship and Regional Development* 32: 197–220. [CrossRef]

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep Learning. *Nature* 521: 436–44. [CrossRef] [PubMed]

Malisetti, Harita, and Punam Singh. 2024. Gender Inequality and Entrepreneurial Choice: Untangling the Roots for Sustainable Growth. *Journal of Entrepreneurship in Emerging Economies* 17: 928–54. [CrossRef]

McVean, Gil. 2009. A Genealogical Interpretation of Principal Components Analysis. *PLoS Genetics* 5: e1000686. [CrossRef] [PubMed]

Molnar, Christoph. 2020. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. Morrisville: Lulu Press.

Nisak, Seli Sofiatun, and Lilik Sugiharti. 2020. Gender Inequality and Women Poverty in Indonesia. *International Journal of Innovation, Creativity and Change* 11: 375–87.

Ogbari, Mercy Ejovwokeoghene, Flourish Folorunso, Busola Simon-Ilogho, Olufunke Adebayo, Kofoworola Olanrewaju, Joy Efegbudu, and Michael Omoregbe. 2024. Social Empowerment and Its Effect on Poverty Alleviation for Sustainable Development among Women Entrepreneurs in the Nigerian Agricultural Sector. *Sustainability* 16: 2225. [CrossRef]

Ogujiuba, Kanayo, Ebenezer Olamide, Isaac Agholor, and Estelle Boshoff. 2021. Are There Gender Differences in Sustainable Entrepreneurship Indicators amongst SMEs in South Africa? Application of MANOVA. *International Journal of Financial Research* 12: 151–63. [CrossRef]

Pueyo, Ana, and Mar Maestre. 2019. Linking Energy Access, Gender and Poverty: A Review of the Literature on Productive Uses of Energy. *Energy Research and Social Science* 53: 170–81. [CrossRef]

Raman, Raghu, Nava Subramaniam, Vinith Kumar Nair, Avinash Shivdas, Krishnashree Achuthan, and Prema Nedungadi. 2022. Women Entrepreneurship and Sustainable Development: Bibliometric Analysis and Emerging Research Trends. *Sustainability* 14: 9160. [CrossRef]

Rousseeuw, Peter J. 1987. Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Journal of Computational and Applied Mathematics* 20: 53–65. [CrossRef]

Roustaei, Narges. 2024. Application and Interpretation of Linear-Regression Analysis. *Medical Hypothesis, Discovery and Innovation in Ophthalmology* 13: 151–59. [CrossRef] [PubMed]

Santos Silva, Manuel, and Stephan Klasen. 2021. Gender Inequality as a Barrier to Economic Growth: A Review of the Theoretical Literature. *Review of Economics of the Household* 19: 581–614. [CrossRef]

Setini, Made, Ni Nyoman Kerti Yasa, I Wayan Gede Supartha, I Gusti Ayu Ketut Giantari, and Ismi Rajiani. 2020. The Passway of Women Entrepreneurship: Starting from Social Capital with Open Innovation, through to Knowledge Sharing and Innovative Performance. *Journal of Open Innovation: Technology, Market, and Complexity* 6: 25. [CrossRef]

Solodoha, Eliran, Stav Rosenzweig, and Shai Harel. 2026. Do Women Entrepreneurs Influence Their Start-Up's Exit? Entrepreneurial Identity and Firm Exit in High-Tech. *Journal of Business Research* 210: 116109. [CrossRef]

Song, Yu-Chen, Hai-Dong Meng, Michael J. O'Grady, and Gregory M. P. O'Hare. 2010. The Application of Cluster Analysis in Geophysical Data Interpretation. *Computational Geosciences* 14: 263–71. [CrossRef]

Strobl, Carolin, Anne-Laure Boulesteix, Achim Zeileis, and Torsten Hothorn. 2007. Bias in Random Forest Variable Importance Measures: Illustrations, Sources and a Solution. *BMC Bioinformatics* 8: 25. [CrossRef] [PubMed]

Sykes, Alan O. 1993. *An Introduction to Regression Analysis*. Coase-Sandor Institute for Law and Economics Working Paper No. 20. Chicago: University of Chicago Law School.

Varma, Sudhir, and Richard Simon. 2006. Bias in Error Estimation When Using Cross-Validation for Model Selection. *BMC Bioinformatics* 7: 91. [CrossRef] [PubMed]

Veckalne, Regina, and Tatjana Tambovceva. 2023. The Importance of Gender Equality in Promoting Entrepreneurship and Innovation. *Marketing and Management of Innovations* 14: 158–68. [CrossRef]

World Bank. 2023. *Gender Data Portal*. Washington: World Bank. Available online: https://genderdata.worldbank.org (accessed on 29 June 2026).

World Economic Forum. 2023. *Global Gender Gap Report 2023*. Geneva: World Economic Forum.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.