Athens Institute
Working Paper No. 2026-2795-40
4 April 2026

The Long Run Trend in Sovereign State Proliferation: Evidence from 1820 to 2020

Gregory T. Papanikos

No.: 2026-2795-40
Date: 4 April 2026
DOI:
ISSN: 2241-2891

Athens Institute (www.atiner.gr)
2026

Athens Institute's Working Paper Series No. 2026-2795-40, 4 April 2026, pages: 1-26

The Long Run Trend in Sovereign State Proliferation: Evidence from 1820 to 2020[1]

*By Gregory T. Papanikos**

This paper examines the long-run drivers of sovereign state formation using decadal global data from 1820 to 2020. Over this period, the number of sovereign states rose from 36 to 195, with a pronounced acceleration during the decolonization era. I develop a parsimonious framework in which human development—captured by per capita income, literacy, and human rights—reduces the costs of statehood while increasing the benefits of self-determination. Using a stationary principal component index of development (PCI) and decadal observations, I show that, once development is accounted for, the secular time trend in state counts becomes statistically indistinguishable from zero, while a robust post-1960 acceleration remains, consistent with the catalytic role of decolonization and shifting international norms. The findings highlight the joint importance of structural development and episodic political shocks in shaping the expansion of the international state system.

Keywords: *Sovereign state formation; development; decolonization; human capital; human rights; literacy; per capita GDP; modernization; international system.*

JEL Codes: *O10; O15; P16; N40; F54.*

Introduction

The expansion of the international state system—from 36 sovereign entities in 1820 to 195 in 2020 (see Figure 1)—represents a fundamental reorganization of global political life.[1] Explanations for this transformation emphasize different mechanisms: geopolitical shocks (wars, imperial collapse), the diffusion of national identity and self-determination norms, and economic modernization that builds administrative capacity and political demand. Each perspective offers valuable insight, but few studies test these ideas using long-run, aggregated global data. In this study, the unit of analysis is the world, and the temporal unit is the decade, reflecting data availability.

*President, Athens Institute & Professor (adjunct), University of Tennessee, Knoxville, USA.

[1]The concept of small, independent political communities exercising self-governance is far older than the modern state system. The ancient Greek world of the Archaic and Classical periods (c. 650–325 BC) was divided into some 1,500 poleis — city-states each consisting of an urban center and its immediate hinterland (Hansen and Nielsen, 2004). The present paper, by contrast, begins from 1820 — the point at which the modern system of territorially bounded, internationally recognized sovereign states, as understood in contemporary international law and political science, can be said to have taken coherent form. The ancient polis tradition underscores that political fragmentation and the demand for self-governance are recurrent features of human organization, but the institutional and normative apparatus of modern sovereignty — standing armies, formal recognition, treaty-based international order — makes the post-1820 period a meaningfully distinct unit of analysis.

This paper tests a simple, testable hypothesis: the long-run increase in the number of sovereign states primarily reflects improvements in human development rather than an autonomous temporal process, with "shocks" such as decolonization acting as catalysts that alter timing and pace. I proceed in three steps. First, I document the decadal evolution of state counts and identify a structural break around the decolonization era. Second, I present a compact theoretical framework linking development to the costs and benefits of independence. Third, I estimate reduced-form models that combine a stationary development index with a post-1960 structural break to separate long-run developmental effects from episodic political accelerations. The results show that development explains the rise in sovereign nation-states, while the post-1960 acceleration remains an independent and quantitatively important effect.

Existing explanations fall into three broad categories. First, political and historical accounts emphasize the role of war, imperial collapse, and geopolitical restructuring in generating new states (Tilly, 1990; Wimmer & Feinstein, 2010; Papanikos, 2025a; Goldstein & Pevehouse, 2017; Coggins, 2011). Second, constructivist approaches highlight the importance of national identity, ideology, and the diffusion of self-determination norms (Anderson, 1983/2006; Reus-Smit, 2001; Papanikos, 2024). Third, economic theories—particularly those associated with the modernization tradition—argue that development generates the conditions necessary for statehood by increasing administrative capacity, facilitating coordination, and strengthening demands for political representation (Lipset, 1959; Rostow, 1960; Alesina & Spolaore, 1997; Alesina, Spolaore & Wacziarg, 2000; Acemoglu et al., 2008; 2009). While each of these perspectives provides valuable insights, they also face limitations: political and historical accounts often explain discrete episodes of state formation but not the persistent upward trend in the number of states; constructivist approaches identify key mechanisms but are difficult to quantify over long time horizons; and economic theories offer a unifying framework but have not been tested using long-run global data.

Quantitative research on secession and state fragmentation (Sorens, 2005; Cederman, Wimmer & Min, 2010; Hale, 2008) highlights the structural conditions under which political units break apart, complementing the long-run global perspective adopted here. A parallel literature emphasizes the role of international recognition norms and the evolving authority of international organizations (Coggins, 2011; Zacher, 2001; Jackson, 1990), providing theoretical support for treating the post-1960 period as a distinct regime shaped by decolonization and changing international rules. Finally, work on the microfoundations of nationalism and self-determination (Gellner, 1983; Laitin, 2007; Wimmer, 2018) underscores how rising income, literacy, and rights expand the demand for sovereign statehood, consistent with the mechanisms formalized in this paper.

The empirical analysis yields three main findings. First, a simple time trend explains a large share of the increase in the number of sovereign states, with a significant acceleration after 1960. Second, once a composite measure of human development is included, the time trend becomes statistically insignificant, indicating that development accounts for the long-run expansion of the state system. Third, the post-1960 acceleration remains significant even after controlling for development,

suggesting that decolonization and changes in international norms played an independent catalytic role.

These findings contribute to several strands of the literature. They provide long-run empirical support for the modernization hypothesis, which links economic and social development to changes in political organization. They complement territory-level studies of state formation by focusing on aggregate global dynamics. Finally, they highlight the interaction between structural economic forces and international political processes in shaping the evolution of the international system.

The analysis is subject to important limitations. The use of long-run, aggregate global data necessarily entails a small number of observations and raises concerns about non-stationarity and spurious correlation. While I address these issues through the construction of stationary indices and the incorporation of structural breaks, the results should be interpreted as documenting robust long-run relationships rather than establishing definitive causal effects. Nonetheless, the consistency of the findings across specifications suggests that development plays a central role in explaining the historical expansion of sovereign states.

The remainder of the paper is organized as follows. Section 2 presents a simple theoretical framework linking development to state formation. Section 3 describes the empirical strategy. Section 4 presents and discusses the empirical results. Section 5 concludes.

Conceptual Framework

This section develops a simple framework linking human development to the formation of sovereign states. The central idea is that the decision to form an independent state depends on a trade-off between the costs and benefits of sovereignty. By "costs and benefits," I refer to all types and categories of costs and benefits as perceived by the citizens of a specific geographical area. These may include economic, political, social, and cultural costs and benefits. While the concept of "sovereignty" is often associated with wars, this is not always the case. A free and independent nation may voluntarily choose to relinquish its sovereignty if the perceived benefits exceed the costs. For example, members of the eurozone have voluntarily given up their sovereign right to maintain their own currency because they judged that a common currency would be more advantageous. Similarly, the peaceful dissolution of Czechoslovakia into the Czech Republic and Slovakia demonstrates that states can choose to secede voluntarily when the perceived benefits of independence outweigh the costs.

Consider a population that can either remain part of a larger political unit or form an independent state. Let the net benefit of sovereignty (B_t) be given by:

$$B_t = U_t - C_t \tag{1}$$

where U_t represents the benefits of self-governance and C_t the costs of establishing and maintaining an independent state. A sovereign state emerges when $B_t > 0$.

Human development affects both components of this expression. The costs of a sovereign nation-state include the provision of public goods, the establishment of administrative institutions, and the capacity to defend territorial integrity. These costs are decreasing in the level of development. First, higher per capita income increases the fiscal capacity of a population, making it easier to finance the fixed costs of governance. Second, education and human capital improve administrative efficiency and reduce coordination costs. Third, technological progress—particularly in communication and transportation—lowers the effective cost of governing a given territory (Branch, 2011). Formally, we can write:

$$C_t = C(D_t),\ \partial C/\partial D_t<0 \quad (2)$$

where D_t is a measure of development.

The historical evidence on this mechanism is compelling. Stasavage (2010) demonstrates that in an era of costly communications and transport, the intensive form of political representation necessary to sustain viable governance — and with it, the capacity to raise revenue through taxation and access public credit — was far easier to maintain in geographically compact polities. When travel was prohibitively costly, it was nearly impossible for constituents to monitor the actions of their representatives, making effective self-governance structurally dependent on small territorial scale. This implies that the threshold level of development required to render a sovereign state viable is not constant: it is inversely related to the geographic and demographic scale of the prospective state, and decreasing in the quality of communication and transport technology. As development and technology improved over the nineteenth and twentieth centuries — the very period captured in this paper's dataset — the effective cost of governing larger territories fell, progressively widening the set of populations for whom independence became feasible. This dynamic is precisely what equation (2) formalizes. As societies become richer and more educated, individuals place greater value on political participation, representation, and the protection of rights. This increases the demand for self-determination.

Human rights and political freedoms play a central role in this process. Greater respect for civil and political rights enables collective action, facilitates the organization of independence movements, and increases the legitimacy of claims to sovereignty in the international system. Thus:

$$U_t = U(D_t),\ \partial U/\partial Dt>0 \quad (3)$$

At the global level, the number of sovereign states can be interpreted as the equilibrium outcome of many such decisions. As development increases over time, more populations satisfy the condition $B_t > 0$, leading to an expansion in the number of independent states. This yields the key empirical prediction:

Prediction 1: Higher levels of human development are associated with a greater number of sovereign states.

However, development alone does not determine the timing of state formation. Political shocks—such as the collapse of empires, wars, or changes in international norms—can alter the constraints faced by potential states. To capture this, let the number of states (S_t) be given by:

$$S_t = f(D_t, Z_t) \quad (4)$$

where Z_t represents political and institutional factors, including the international environment, while *f* represents the functional form, as discussed below in the empirical section of the paper.

In particular, the post-1960 decolonization period can be interpreted as a shift in Z_t that reduced barriers to state formation by legitimizing self-determination and facilitating international recognition. This leads to a second prediction:

Prediction 2: Political shocks that relax constraints on sovereignty increase the rate of state formation, conditional on the level of development.

The framework highlights two complementary forces driving the expansion of the state system. Development generates the underlying capacity and demand for independence, while political and institutional changes determine when and how this potential is realized. The empirical analysis that follows tests these predictions using long-run global data.

Empirical Strategy

Data

The dependent variable is the total number of sovereign states, S_t, worldwide, measured decennially from 1820 to 2020, based on data availability. Observations are recorded according to current country borders, following Butcher and Griffiths (2020). The dataset includes 36 states in 1820, expanding to 195 states by 2020 (see Figure 1). The main explanatory variables capturing human development are:

1. Per capita GDP (Y/N).
2. Literacy – Share of adults who can read and write a simple statement about their everyday life.
3. Human Rights Index (HR) – based on V-Dem expert estimates, including political freedoms and property rights.

These variables are combined into a principal component index (PCI) to summarize the multifaceted concept of human development.

Control variables include:

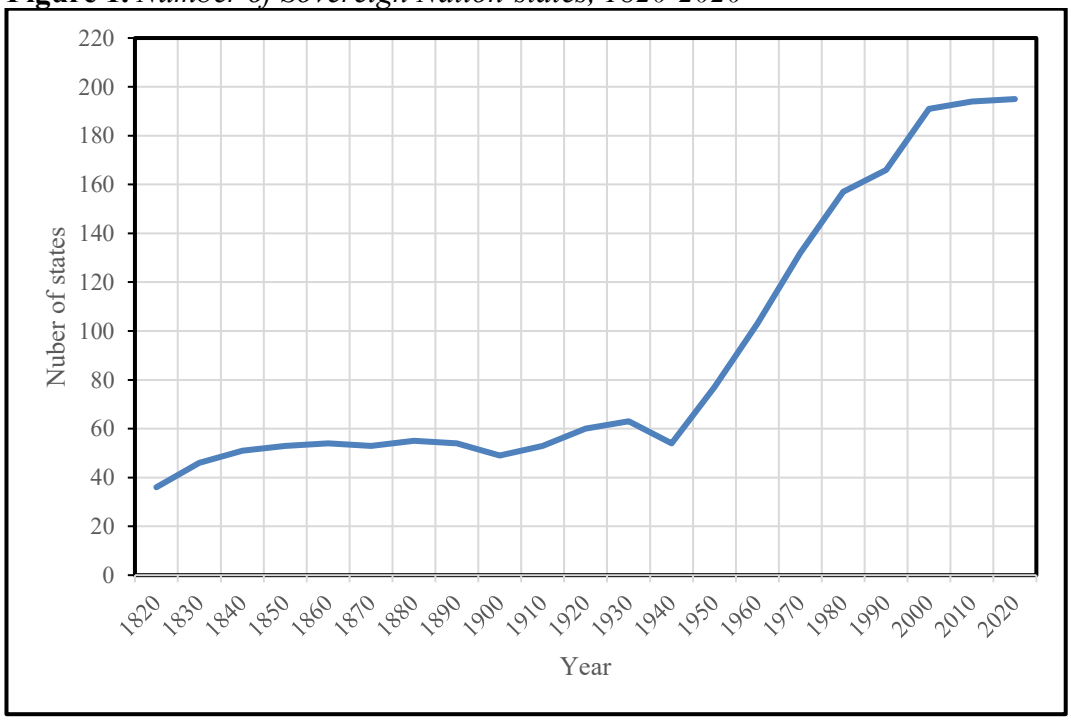
1. Trade openness (TO): (exports + imports)/GDP
2. Population (N)
3. Political regime type (fraction of population living in autocracies)

All data sources are publicly available and cover the full 1820–2020 period, with careful adjustments made for missing values and interpolation where necessary. The sources and raw data are provided in Appendix A and Appendix B, respectively. Table 1 presents summary statistics for all variables—both statistically significant and non-significant—used in the empirical analysis. Examining the summary statistics for this panel of 21 observations, several important patterns emerge across the ten variables. A brief description of each variable is provided below.

Number of Sovereign States (S)

The sovereign state count has a mean of 90 and a median of only 55, a gap of 35 that immediately signals a right-skewed distribution: for the majority of the sample period the world contained far fewer states than the late-twentieth-century surge would suggest. The range — from 36 in 1820 to 195 in 2020 — captures the full arc of modern state proliferation, a more than fivefold increase over two centuries. The standard deviation of 56 is large relative to the mean (coefficient of variation of 62%), reflecting the uneven pace of expansion: slow and gradual before decolonization, rapid and concentrated after 1960. The positive skewness of 0.96 and the below-normal kurtosis of 2.23 together indicate a distribution stretched to the right by the post-1960 acceleration, but without an extreme outlier driving the tail — the expansion was sustained rather than a single jump. The Jarque-Bera statistic of 3.71 with a p-value of 0.16 does not reject normality at conventional levels, meaning that while the distribution is not perfectly symmetric, a normal approximation is not unreasonable for regression purposes. The use of ln(S) as the dependent variable in the empirical models helps stabilize this mild asymmetry.

Figure 1. *Number of Sovereign Nation-states, 1820-2020*



World GDP in billions of 2021 international USD (Y)

World GDP is the most non-normal variable in the dataset by a wide margin. The mean of approximately $26,700 billion sits far above the median of $6,684 billion — a gap of over $20,000 billion — reflecting the explosive compounding of global output in the second half of the twentieth century. The maximum of $147,693 billion is nearly ninety times the minimum of $1,628 billion, and the extreme positive skewness of 1.86 combined with a kurtosis of 5.43 — well above the normal benchmark of 3 — is a near-certain signature of exponential growth. The Jarque-Bera statistic of 17.3 decisively rejects normality ($p = 0.0002$). The coefficient of variation of 152% further underscores that the mean is a poor summary of central tendency for this series. In practical terms, Y in levels is almost certainly a non-stationary, and its inclusion in any regression in levels would risk severe spurious correlation. This is precisely why the empirical strategy constructs a stationary composite index (PCI) rather than using Y directly, and why log transformation or first-differencing would be essential if Y were used as a regressor.

World Population in billions (N)

World population grows from 1.09 billion in 1820 to 7.89 billion in 2020, a more than sevenfold increase that reflects the global demographic transition. The mean of 2.86 billion and median of 1.90 billion diverge by nearly one billion, again pointing to right skewness driven by the acceleration of population growth in the post-war period. The standard deviation of 2.09 billion is large relative to the mean (CV of 73%), and the kurtosis of 3.18 is close to normal, indicating that the tail behavior is roughly Gaussian even if the distribution is asymmetric. The Jarque-Bera p-value of 0.07 sits just outside the 5% rejection region, suggesting approximate non-normality. Like Y, population in levels is a strong candidate for a unit root process — the stationarity tests in Table 2 confirm this — and should not be used as a regressor in levels. Its role here is primarily as a denominator in the construction of per-capita GDP (YN).

Per Capita GDP in 2021 international USD (YN)

Per capita income rises from $1,494 in 1820 to $18,726 in 2020, a twelvefold increase that encapsulates the long-run process of economic modernization at the heart of the paper's theoretical framework. The mean of $5,815 exceeds the median of $3,527 by $2,288, and the skewness of 1.28 and kurtosis of 3.49 reflect the compounding of both output growth and population change over time. The coefficient of variation of 87% indicates very high dispersion relative to the mean, and the Jarque-Bera p-value of 0.05 sits precisely at the 5% threshold, making normality a borderline assumption. As with Y and N, per-capita GDP is a trending series and Table 2 classifies it as I(2) in levels and I(1) in logs, confirming it must be transformed or replaced by a stationary composite for regression use. Its contribution to long-run development is captured in the PCI index, which pools YN with literacy and human rights into a stationary linear combination.

Trade Openness as a % of GDP (TO)

Trade openness ranges from 9.5% in 1820 to a peak of 56.3% in 2010, before retreating slightly to 51.7% in 2020—a pattern consistent with the well-documented waves of globalization and deglobalization over the past two centuries. The mean (22.6%) and median (17.0%) diverge moderately, resulting in a positive skewness of 1.14 that reflects the sharp increase in trade integration during the final decades of the sample. Importantly, the kurtosis of 2.91 is close to 3, and the Jarque-Bera p-value of 0.10 just fails to reject normality. The coefficient of variation of 66% is notable but far smaller than those of Y, N, and YN. Despite its relative distributional stability, Table 2 classifies TO as I(1), indicating that it carries a stochastic trend and must be treated accordingly in regression analysis. It enters the empirical models as a control variable rather than as a primary driver.

Education as a % of Adults with Formal Schooling (E)

Education rises from 17.2% in 1820 to 86.9% in 2020, spanning virtually the entire range of the variable's natural bounds. The mean of 45.7% and median of 38.6% are relatively close together (a gap of only 7 percentage points), and the skewness of 0.46 is the second lowest in the dataset after DE (in absolute terms), indicating a broadly symmetric distribution with a modest right tail. The kurtosis of 1.73 is well below 3, pointing to a platykurtic (flat-topped) distribution — observations are spread fairly evenly across the range rather than concentrated around the mean. The Jarque-Bera p-value of 0.34 is the second highest in the dataset, providing no grounds to reject normality. Despite these favorable distributional properties, Table 2 classifies E as I(2) — the most strongly non-stationary series in the dataset — because the level rises steadily and accelerates, with no reversion. Education is therefore excluded from direct use as a regressor and its long-run information is instead captured through the PCI index.

Literacy Rate as a % of Adults (L)

Literacy increases from 12.1% in 1820 to 86.8% in 2020, closely tracking the expansion of education but starting from a lower base and converging more slowly. The mean (38.7%) and median (31.6%) diverge by about 7 percentage points, mirroring the pattern in EEE, and the skewness of 0.78 is moderately positive. The kurtosis of 2.13 is below the normal value, indicating a relatively flat distribution across the historical range. The Jarque-Bera p-value of 0.25 does not reject normality. Notably, the correlation between L and E over the sample is very high, as both capture the accumulation of human capital. The paper uses L (rather than E) as one of the three components of PCI because literacy has broader historical coverage and is classified as I(1) rather than I(2), making it more suitable for inclusion in a stationary linear combination with YN and HR. The close distributional similarity between E and L, combined with L's superior stationarity properties, supports this choice. However, when E is used instead of L in constructing PCI, the results remain very close.

Human Rights Index from 0 to 1 (HR)

The human rights index is the most tightly bounded variable in the dataset, ranging from 0.333 in 1820 to 0.700 in 2020, with a mean of 0.46 and a standard deviation of only 0.11 — giving the lowest coefficient of variation (23%) of any variable. This compression reflects the bounded nature of the index and the fact that global human rights conditions, while improving over two centuries, have never approached the theoretical maximum of 1 within the sample period. The median of 0.43 is close to the mean, but the positive skewness of 1.27 and kurtosis of 3.46 — slightly above the normal benchmark — indicate a right-skewed distribution with modest excess peakedness: most observations cluster in the lower portion of the range, with the higher values concentrated in the most recent decades. The Jarque-Bera p-value of 0.055 sits just above the 5% threshold, meaning normality is a borderline but broadly defensible working assumption. Table 2 classifies HR as I(0) with a structural break, confirming it is stationary in levels and suitable for direct inclusion in regression models — a key property that motivates its role both as a component of PCI and as a standalone regressor in Model 3.

Autocracy Share measured as a Fraction of Population in Autocracies (DE)

The autocracy variable has a mean of 0.63 and a median of 0.75 — here the median *exceeds* the mean, producing the only negatively skewed variable in the dataset (skewness of −0.69). This reversal is economically meaningful: for most of the sample period the majority of the world's population lived under autocratic rule, with democratic governance concentrated in the most recent decades pulling the mean downward relative to the median. The range spans from 0.23 (2010 and 2020) to 0.86 (1840), reflecting the historically exceptional nature of the recent democratic expansion. The kurtosis of 1.99 is well below 3, indicating a platykurtic distribution with a long lower tail as democratic governance has spread. The Jarque-Bera p-value of 0.28 does not reject normality. The standard deviation of 0.22 produces a coefficient of variation of 35% — the second lowest in the dataset after HR — reflecting that despite the secular downward trend in autocracy, the distribution is relatively compressed. DE enters the empirical models as a control variable; its classification as I(1) in Table 2 means it is included with appropriate caution.

Table 1. *Summary Statistics*

	S	Y	N	YN	TO	E	L	HR	DE	PCI
Mean	90	26697	2.86	5814.6	22.6	45.66	38.73	0.46	0.63	0
Median	55	6684	1.9	3527	16.6	38.6	31.62	0.43	0.75	-0.33
Maximum	195	147693	7.9	18726	56.3	86.9	86.75	0.7	0.86	3.15
Minimum	36	1628	1.1	1494	9.5	17.2	12.05	0.33	0.23	-2.46
Std. Dev.	56	40567	2.1	5073	14.82	24.25	25.63	0.11	0.22	1.71
Skewness	0.96	1.86	1.22	1.28	1.14	0.46	0.78	1.27	-0.69	0.57
Kurtosis	2.22	5.43	3.18	3.49	2.9	1.73	2.13	3.46	1.99	2.27
Jarque-Bera	3.71	17.3	5.3	5.9	4.6	2.16	2.8	5.8	2.6	1.6

	S	Y	N	YN	TO	E	L	HR	DE	PCI
Probability	0.16	0.0002	0.07	0.05	0.1	0.34	0.25	0.05	0.28	0.47
Observations	21	21	21	21	21	21	21	21	21	21

Note: See appendix A for the sources of data and Appendix B for the raw data

Principal Component Index (PCI)

By construction, the PCI has a mean of zero and is built from standardized inputs (per-capita GDP, literacy, and human rights), so its mean and scale are artefacts of normalization rather than substantive statistics. The standard deviation of 1.71 and range from −2.46 in 1820 to 3.15 in 2020 reflect the cumulative improvement in human development over the sample: the index rises by nearly 5.6 standard units over two centuries, capturing the co-movement of income, education, and rights. The positive skewness of 0.57 indicates a mild right tail, consistent with the acceleration of development in the late twentieth century pulling a few high observations above the central tendency. The kurtosis of 2.28 is slightly below 3, suggesting a distribution somewhat flatter than normal. The Jarque-Bera p-value of 0.45 is the highest in the dataset, providing no grounds whatsoever to reject normality — the PCI is, from a distributional standpoint, the most regression-friendly variable in the table. Critically, Table 2 classifies PCI as I(0) with a structural break, confirming stationarity in levels. This property is the primary motivation for constructing the index: it summarizes the long-run development trend without inheriting the explosive non-stationarity of its individual components, thereby enabling its inclusion in OLS regressions without risking spurious results.

Trending Variables and Non-stationarity

Many explanatory variables, including GDP, population, and education, exhibit strong secular trends. Regressing these trending series on a trending dependent variable risks spurious correlation. To address this:

1. I test for unit roots using breakpoint-augmented Dickey-Fuller tests (Perron, 1989), allowing for structural breaks in 1960 (the onset of decolonization). The findings are reported in Table 2.
2. To capture the long-run co-movement of income, human capital, and institutional quality, I construct a Principal Component Index (PCI) using per-capita GDP, literacy, and human rights. After standardizing each variable, principal component analysis yields a dominant first component that explains 94.7% of the total variance, with nearly equal positive loadings on all three inputs. The resulting PCI provides a parsimonious, stationary summary of global human development suitable for inclusion in the empirical models (see Appendix C). As a robustness check, I construct an alternative PCI by replacing literacy with the share of the world population living under autocracy (DE). The resulting index is extremely similar to the baseline measure, with a correlation of approximately 0.93 over the 1820–2020 period. PCI again yields a dominant first component explaining more than 93% of the variance, with positive loadings on income and human rights and a

negative loading on autocracy. This confirms that the long-run developmental trend captured by the PCI is robust to alternative measures of human development and political modernization.

Table 2. *Stationary Properties*

Variable	Stationarity (levels)	Stationarity (logs)
S	I(0)*	I(0)*
Y	I(1)	I(1)
N	I(1)*	I(1)*
YN	I(2)*	I(1)
TO	I(1)	I(1)
E	I(2)	I(2)
L	I(1)*	I(1)*
HR	I(0)*	I(1)
DE	I(1)	I(1)
PCI	I(0)*	
*Break point unit root test		

The dependent variable (S) is stationary in both levels and logs once the 1960 structural break is accounted for. This is an important and reassuring result. A non-stationary dependent variable would require either differencing — which would change the research question from explaining the level of sovereign state counts to explaining their rate of change — or a cointegration framework. The finding that ln(S) is I(0)* means that OLS regression in levels is a valid strategy, provided the regressors are also stationary or are combined into a stationary index. It also confirms that the visual impression from Figure 1 of a permanently accelerating series is primarily a structural break phenomenon rather than a stochastic trend: once the post-1960 shift is modelled explicitly through the slope dummy interaction, the residual series reverts to a stable mean.

Reading Table 2 as a whole, three conclusions stand out. First, the structural break at 1960 is pervasive: six of the ten variables require the breakpoint-augmented test, confirming that the decolonization era represents a genuine discontinuity in the data-generating processes of development, demography, and state formation simultaneously. Second, the log transformation is beneficial but insufficient for the most non-stationary variables: it reduces YN from I(2)* to I(1), which is helpful, but leaves E at I(2) and leaves most I(1) variables unchanged. Third, the PCI construction is vindicated by the stationarity results: the only variable in the dataset that is both stationary and rich in development content is the composite index, making it the natural and methodologically justified regressor for the central hypothesis test.

Structural Breaks and Political Shocks

The post-1960 decolonization wave represents a clear structural shock. To model this, I define a dummy interaction $D_t \cdot t$ capturing the change in slope of state formation after 1960:

$$D_t = \begin{cases} 0 \ if \ t < 1960 \\ 1 \ if \ t \geq 1960 \end{cases}$$

This allows the estimation to distinguish between long-run development trends and episodic political accelerations.

Baseline Model

The baseline specification is:

(5) $$\ln(N_t) = \beta_0 + \beta_1 t + \beta_2 (D_t \cdot t) + \varepsilon_t$$

Where:

- t is the decadal time index
- $D_t \cdot t$ captures the acceleration of state formation after 1960
- ε_t is the error term

This specification provides a descriptive benchmark, documenting the temporal pattern of state formation without yet incorporating development.

Modernization Model

To test the central hypothesis that development drives state formation, I augment the baseline with the stationary principal component index (PCI):

(6) $$\ln(N_t) = \beta_0 + \beta_1 t + \beta_2 (D_t \cdot t) + \beta_3 PCI + \varepsilon_t$$

Interpretation:

- $\beta 3 > 0$ indicates that higher human development is associated with more sovereign states.
- If $\beta 1 \approx 0$ after including PCI, this suggests that the apparent time trend is explained by development, consistent with the modernization hypothesis.

Human Rights Channel

To investigate the mechanism linking development and sovereignty, I replace PCI with HR:

$$\ln(N_t) = \beta_0 + \beta_1 t + \beta_2 (D_t \cdot t) + \beta_3 HR + \varepsilon_t \quad (7)$$

This specification isolates the political and institutional dimension of development, separating it from the economic and educational components.

Estimation and Inference

Given the small number of observations (21 decadal periods):

- I use OLS with Newey-West HAC standard errors to account for potential autocorrelation.
- The dependent variable is log-transformed ln(S) to stabilize variance and interpret coefficients as approximate growth rates.
- Stationarity tests confirm that the PCI and HR series are suitable for inclusion in levels, mitigating concerns about spurious correlation.

Durbin-Watson statistics and visual inspection of residuals indicate no remaining autocorrelation.

Identification Strategy

While the analysis is primarily descriptive and correlational, several design features strengthen causal inference:

1. Stationary index construction addresses trending variables.
2. Post-1960 structural break exploits quasi-experimental variation due to decolonization.
3. Robustness checks (not shown here) include alternative breakpoints.

The combination of these strategies allows the estimation to separate long-run developmental effects from political shocks, providing credible evidence on the drivers of sovereign state formation.

Empirical Predictions

Based on the framework in Section 2 we make the following predictions:

1. Prediction 1 (Development): Higher levels of human development are associated with a greater number of sovereign states.
2. Prediction 2 (Political Shock): Political shocks that relax sovereignty constraints (e.g., decolonization) accelerate state formation conditional on development.

All claims are presented cautiously, highlighting associations rather than strict causation. The empirical design integrates macro-trends and political shocks while remaining transparent about data limitations.

Estimation Results

Table 3 presents the estimation results of the three empirical models, reporting only specifications whose coefficients are statistically significant. Three models are shown, and the following discussion briefly summarizes the empirical findings for each specification.

A potential concern is the endogeneity of development in explaining state formation. While the theoretical framework treats human development as a driver of sovereignty demand, the causal relationship may run in both directions: the creation of a sovereign state can itself promote economic growth, expand literacy, and improve rights protections. In addition, unobserved geopolitical cycles—such as wars, imperial expansion, or shifts in international norms—may simultaneously influence both development and the number of states. Therefore, the empirical results should be interpreted as documenting robust long-run associations rather than strict causal effects. The use of a stationary composite index (PCI) and a theoretically motivated post-1960 structural break mitigates, but does not eliminate, these concerns.

Model 1: Baseline Model

The baseline model regresses ln(S) on a linear time trend and a post-1960 slope interaction, with no development variables. It establishes the descriptive temporal pattern of sovereign state formation before any economic or institutional explanation is introduced.

The constant of 3.77 (SE = 0.07, t = 53.9) is precisely estimated and implies that at the origin of the time index the log of the sovereign state count is approximately 3.77, corresponding to roughly 43 states — consistent with the historical record at the beginning of the nineteenth century. The linear trend coefficient of 0.032 (SE = 0.01, t = 3.2) is statistically significant at 1% and implies that the number of sovereign states grew at approximately 3.2% per decade throughout the pre-1960 period, reflecting the slow erosion of imperial structures and the gradual emergence of independent polities from the early nineteenth century through decolonization's eve.

The post-1960 interaction term ($D_t \cdot t$) is 0.045 (SE = 0.006), and the note reports a HAC-corrected t-statistic of 7.9, making it by far the most precisely estimated coefficient in the table. This term captures the additional slope in state formation after 1960. Adding the two trend components, the implied post-1960 growth rate is approximately 8.0% per decade — more than double the pre-1960 pace — which is fully consistent with the historical record of rapid decolonization and UN membership expansion between 1960 and 1980. The Durbin-Watson statistic of 1.96 is very close to 2, providing no indication of residual serial correlation and validating the OLS inference. The R^2 of 0.9665 and adjusted R^2 of 0.9628 indicate that two-time variables alone explain over 96% of the variation in the log sovereign state count — a high share that reflects the strongly trending nature of the dependent variable and sets the benchmark against which the development models must be evaluated.

Table 3. *Estimation Results*

Variable	**Baseline model** (1)	**Modernization Model** (2)	**Human Rights Model** (3)
Constant	3.77*** (0.07)	4.15*** (0.04)	3.49*** (0.15)
Trend (t)	0.032*** (0.01)	—	0.028*** (0.007)
Trend·Dummy (t·D)	0.045*** (0.006)	0.034*** (0.01)	0.039*** (0.01)
Principal Component Index (PCI)	—	0.166*** (0.007)	—
Human Rights (HR)	—	—	0.76** (0.34)
R^2	0.9665	0.9749	0.9724
Adjusted R^2	0.9628	0.9721	0.9675
F-statistic (probability)	260 (0.0000)	349 (0.0000)	199 (0.0000)
Durbin-Watson	1.96	1.86	1.90

*Note: standard errors in parentheses below each coefficient. *** significant at 1%, ** significant at 5%. HAC standard errors confirm OLS inference in both specifications. t-statistic on t·D is 7.9 in Model 1. t-statistic on HR is 2.21 in Model 3.*

Model 2: Modernization Model

The modernization model augments the baseline by adding the stationary Principal Component Index (PCI) and dropping the linear time trend, which is the central test of the paper's main hypothesis: that human development, rather than the passage of time itself, drives the long-run expansion of the state system.

The PCI coefficient of 0.166 (SE = 0.007, t = 23.7) is positive, highly significant at 1%, and by far the most precisely estimated substantive coefficient across all three models. A one standard deviation increase in PCI (1.71 units) is associated with an increase in the number of sovereign states of approximately 32.8%, which is an economically large and meaningful effect. The interpretation is exactly as the modernization hypothesis predicts: higher composite human development — combining per-capita income, literacy, and human rights — is robustly associated with a greater number of sovereign entities in the international system.

Critically, the linear time trend is omitted from this specification because it was found to be statistically insignificant. This reflects the finding that, once PCI is included, the time trend becomes statistically indistinguishable from zero and is therefore appropriately excluded—its explanatory content being absorbed by the development index. This constitutes the paper's central result: the secular increase in sovereign states is not an autonomous temporal process but rather a reflection of the underlying trajectory of human development.

The post-1960 interaction term remains at 0.034 (SE = 0.01, t = 3.4), significant at 1%, but its magnitude has fallen by approximately 24% relative to the baseline (from 0.045 to 0.034). This reduction indicates that part of what appeared in Model 1 as a pure post-1960 political acceleration is in fact attributable to the unusually rapid development gains of the postwar decades — but a substantial, independent decolonization effect survives. The improvement in model fit is meaningful: R^2 rises to 0.9749 and the adjusted R^2 to 0.9721, both above Model 1 despite the loss of the trend variable, confirming that PCI is a more parsimonious and informative predictor than time alone. The F-statistic of 349 is the highest across all three models. The Durbin-Watson of 1.86 remains in the acceptable range, supporting the absence of residual autocorrelation.

Model 3: Human Rights Model

The human rights model replaces PCI with the HR index alone, isolating the political and institutional dimension of development from its economic and educational components. This specification tests whether rights improvements independently predict state formation, and whether the time trend re-emerges once the full development composite is disaggregated.

The HR coefficient of 0.76 (SE = 0.34, t = 2.24, reported in the note as 2.21 after rounding) is positive and significant at 5%, confirming that improvements in civil and political rights are associated with a greater number of sovereign states. The economic magnitude is substantial when evaluated over the empirical range of the HR variable: across the full observed variation of 0.367 index points (from 0.333 in 1820 to 0.700 in 2020), the implied increase in the number of states is approximately 32%, closely comparable to the one-standard-deviation effect of the full PCI in Model 2. This confirms that the human rights channel is not merely a proxy for income or education — it carries independent and quantitatively meaningful explanatory power.

An important and informative difference from Model 2 is that the linear time trend re-enters and remains significant in Model 3: the trend coefficient of 0.028 (SE = 0.007, t = 4.0) is significant at 1%. This implies a pre-1960 growth rate of approximately 2.8% per decade, somewhat lower than in the baseline but still precisely estimated. The re-emergence of the trend is theoretically revealing: HR captures the political and institutional dimension of development but does not absorb the economic and literacy components. The residual time trend therefore reflects the additional development content — income growth and literacy expansion — that HR alone cannot proxy. In other words, Model 3 provides a partial test of the modernization hypothesis, while Model 2 provides the full test.

The post-1960 interaction term of 0.039 (SE = 0.01, t = 3.9) is highly significant and its magnitude falls between those of Models 1 and 2. It has declined by 13% relative to the baseline (compared to 24% in Model 2), which is consistent with HR capturing only part of the development-driven component of the post-1960 acceleration. The R^2 of 0.9724 and adjusted R^2 of 0.9675 are slightly below Model 2 but above Model 1, and the F-statistic of 199 remains high. The Durbin-Watson of 1.90 is satisfactory.

The positive and significant coefficient on the human rights index (HR) in Model 3 is consistent with the broader historical account offered by Donnelly (2014), who traces the co-evolution of sovereignty and human rights norms since 1945. In the immediate postwar decades, sovereignty was constructed as absolute territorial exclusivity — a framework that insulated states from external scrutiny of domestic rights practices and thereby raised the threshold for international recognition of new states. As Donnelly documents, this absolutist conception gradually gave way to a more conditional understanding of sovereignty in which states are expected to meet minimum human rights standards to enjoy full international legitimacy. The empirical results reflect this dynamic at the aggregate level: periods of improving human rights are associated with a greater number of sovereign states, because rising rights standards simultaneously increase the demand for self-determination among populations seeking accountable governance and raise the international legitimacy of independence claims. In this sense, the HR channel identified here captures not merely a domestic political mechanism — greater civil and political freedoms facilitating the organization of independence movements — but also an international one: the progressive alignment of sovereignty with rights performance that Donnelly traces historically, and which the post-1960 decolonization wave both exemplified and accelerated.

Cross-Model Interpretation

Reading across all three models, a coherent and mutually reinforcing picture emerges. The baseline establishes that state formation has a strong temporal structure, with a pre-1960 steady pace and a sharp post-1960 acceleration. The modernization model shows that once composite human development is controlled for, time itself has no independent explanatory power — the apparent trend is simply the footprint of development. The human rights model shows that even a single institutional dimension of development carries significant explanatory power, though it cannot fully displace the time trend the way the full PCI can. Together, the three models support a two-part interpretation of the expansion of the international state system: development generates the long-run structural conditions for sovereign statehood, while decolonization and the post-1945 international order provided the discrete political catalyst that translated latent development potential into realized sovereignty. The consistency of the post-1960 interaction term across all three specifications — ranging narrowly from 0.034 to 0.045 — is particularly reassuring, confirming that the decolonization effect is not an artefact of any particular development measure and is a robust feature of the data.

Relationship to Existing Literature

These findings contribute to several strands of the literature. First, they provide long-run empirical support for the modernization hypothesis by linking development to a fundamental transformation in political organization. While previous work has focused on democracy or institutional quality, this paper extends the analysis to the formation of sovereign states.

Second, the results complement territory-level studies of state formation by shifting the focus to aggregate global dynamics. Rather than explaining why specific regions secede, the analysis addresses why the overall number of states increases over time.

Third, the findings highlight the importance of international norms and institutions, aligning with constructivist accounts of sovereignty while embedding them within a broader development framework.

The findings of this paper also contribute to the ongoing debate on the income–democracy nexus. Acemoglu et al. (2008, 2009) show that, once country fixed effects absorb time-invariant confounders, the cross-sectional correlation between per capita income and democratic institutions largely disappears. Cervellati et al. (2014) refine this conclusion by demonstrating that the pooled null result masks important heterogeneity: income promotes democracy among non-colonies but is associated with lower levels of democracy in former colonies, where colonial institutions and critical junctures mediate the relationship.

The present paper sidesteps the fixed-effects identification problem by working with a single global time series rather than a cross-section of countries, and by constructing a stationary composite index (PCI) that captures the co-movement of income, literacy, and human rights over the long run. In this sense, the approach is complementary to the country-level literature. Whereas Acemoglu et al. (2008, 2009) and Cervellati et al. (2014) ask whether richer countries are more democratic, holding country characteristics constant, this paper asks whether a richer, more literate, and more humanitarian world—characterized by greater respect for human rights—generates more sovereign states, a fundamentally different aggregate question.

Limitations

Several limitations should be noted. First, the analysis relies on a small number of decadal observations, which constrains statistical power and limits the use of more sophisticated time-series techniques. Second, despite efforts to address non-stationarity, the possibility of residual spurious correlation cannot be fully ruled out. Third, the empirical design does not provide a definitive causal identification strategy, and the results should therefore be interpreted as documenting robust long-run associations.

These limitations suggest caution in interpreting the estimates quantitatively, but they do not undermine the qualitative conclusion that development is closely linked to the expansion of sovereign nation-states.

Summary Interpretation

Taken together, the results point to a coherent interpretation of the evolution of the international system. The expansion of sovereign states over the past two centuries reflects a broader process of structural transformation driven by human development. Political shocks—most notably decolonization—accelerated this process by relaxing constraints on sovereignty, but did not fundamentally alter its underlying trajectory.

Conclusions

This paper examines the long-run evolution of the sovereign state system and provides evidence that its expansion is closely linked to the process of human development. Using global data from 1820 to 2020, I show that the steady increase in the number of sovereign states can be largely accounted for by improvements in income, education, and human rights. Once these factors are taken into account, the secular time trend in state formation disappears. At the same time, the post-1960 acceleration in state formation remains robust, reflecting the independent role of decolonization and changes in the international political environment.

These findings suggest a reinterpretation of one of the central transformations of the modern world. The proliferation of sovereign states does not appear to be an autonomous historical process driven solely by geopolitical shocks or ideological change. Instead, it reflects a broader pattern of structural transformation in which development expands both the capacity for and the demand for political independence. Political events such as decolonization are best understood as catalysts that enable this underlying process to manifest.

More broadly, the results highlight the importance of integrating economic and political explanations in the study of institutional change. The evolution of the international system is shaped not only by shifts in power and norms, but also by the gradual accumulation of human and economic capabilities within societies. This perspective helps reconcile competing accounts of state formation by placing them within a unified framework.

At the same time, several questions remain open. The analysis is conducted at a highly aggregated level and abstracts from regional heterogeneity and strategic interactions among states. Future work could exploit panel data or quasi-experimental variation to more precisely identify the causal channels linking development to sovereignty. In addition, the framework developed here may be applied to other forms of political reorganization, including regional integration and state fragmentation.

Understanding the forces that shape the boundaries of political authority remains a central question in economics and political science. This paper provides evidence that development plays a fundamental role in this process, offering a new lens through which to interpret the historical expansion of the state system and its possible future trajectory.

The findings also carry implications for the future of global fragmentation. If the long-run relationship documented here continues to hold, rising levels of income, literacy, and rights may sustain pressures for political self-determination in the twenty-first century. At the same time, the post-1960 shift in international recognition norms suggests that institutional environments will continue to shape whether such demands translate into new sovereign states. While the results do not imply a mechanical prediction of further proliferation, they indicate that the structural drivers associated with past waves of state formation remain present—and in some regions, are intensifying. Understanding how development and international norms interact will therefore be central to anticipating the future trajectory of the global state system.

References

Acemoglu, D., Johnson, S., Robinson, J. A., & Yared, P. (2008). Income and democracy. *American Economic Review, 98*(3), 808–842. https://doi.org/10.1257/aer.98.3.808.

Acemoglu, D., Johnson, S., Robinson, J. A., & Yared, P. (2009). Reevaluating the modernization hypothesis. *Journal of Monetary Economics, 56*(8), 1043–1058. https://doi.org/10.1016/j.jmoneco.2009.10.002.

Acemoglu, D., Naidu, S., Restrepo, P., & Robinson, J. A. (2019). Democracy does cause growth. *Journal of Political Economy, 127*(1), 47–100. https://doi.org/10.1086/700936.

Alesina, A., & Spolaore, E. (1997). On the number and size of nations. *The Quarterly Journal of Economics, 112*(4), 1027–1056. https://doi.org/10.1162/003355305555541.

Alesina, A., Spolaore, E., & Wacziarg, R. (2000). Economic integration and political disintegration. *American Economic Review, 90*(5), 1276–1296. https://doi.org/10.1257/aer.90.5.1276.

Anderson, B. (2006). *Imagined communities: Reflections on the origin and spread of nationalism* (Rev. ed.). Verso. (Original work published 1983)

Branch, J. (2011). Mapping the sovereign state: Technology, authority, and systemic change. *International Organization, 65*(1), 1–36.

Butcher, C., & Griffiths, R. (2020). *International System Dataset version 2* [Data set]. With minor processing by Our World in Data. Retrieved March 18, 2026, from https://archive.ourworldindata.org/20251208-134119/grapher/sovereign-state-butcher-griffiths.html

Cederman, L.-E., Wimmer, A., & Min, B. (2010). Why do ethnic groups rebel? New Data and Analysis. *World Politics, 62*(1), 87–119.

Cervellati, M., Jung, F., Sunde, U., & Vischer, T. (2014). Income and democracy: Comment. *American Economic Review, 104*(2), 707–719. https://doi.org/10.1257/aer.104.2.707

Coggins, B. (2011). Friends in high places: International politics and the emergence of states from secessionism. *International Organization, 65*(3), 433–467.

Donnelly, J. (2014). State sovereignty and international human rights. *Ethics & International Affairs, 28*(2), 225–238.

Gellner, E. (1983). *Nations and nationalism*. Cornell University Press.

Goldstein, J. S., & Pevehouse, J. C. (2017). *International relations* (11th ed.). Pearson.

Hale, H. (2008). *The foundations of ethnic politics: Separatism of states and nations in Eurasia and the world*. Cambridge University Press.

Hansen, M. H., & Nielsen, T. H. (Eds.). (2004). *An inventory of archaic and classical poleis*. Oxford University Press.

Jackson, R. H. (1990). *Quasi-states: Sovereignty, international relations and the Third World*. Cambridge University Press.

Laitin, D. D. (2007). *Nations, states, and violence*. Oxford University Press.

Lipset, S. M. (1959). Some social requisites of democracy: Economic development and political legitimacy. *American Political Science Review, 53*(1), 69–105. https://doi.org/10.2307/1951731

Papanikos, G. T. (2024). The national identity of ancient and modern Greeks. *Athens Journal of Mediterranean Studies, 10*(1), 63–80.

Papanikos, G. T. (2025a). Is the Thucydides trap relevant to the rise and fall of the Ottoman Empire? A cliometric analysis. *Athens Journal of Mediterranean Studies, 11*(2), 1–17.

Papanikos, G. T. (2025b). Are human rights a luxury or a normal good? *Athens Journal of Politics and International Affairs, 1*(1), 9–38.

Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica, 57*(6), 1361–1401.

Reus-Smit, C. (2001). Human rights and the social construction of sovereignty. *Review of International Studies, 27*(4), 519–538.

Rostow, W. W. (1960). *The stages of economic growth: A non-communist manifesto*. Cambridge University Press.

Sorens, J. (2005). The cross-sectional determinants of secessionism in advanced democracies. *Comparative Political Studies, 38*(3), 304–326.

Stasavage, D. (2010). When distance mattered: Geographic scale and the development of European representative assemblies. *American Political Science Review, 104*(4), 625–643.

Tilly, C. (1990). *Coercion, capital, and European states, AD 990–1990*. Blackwell.

Wimmer, A. (2018). *Nation building: Why some countries come together while others fall apart*. Princeton University Press.

Wimmer, A., & Feinstein, Y. (2010). The rise of the nation-state across the world, 1816 to 2001. *American Sociological Review, 75*(5), 764–790.

Zacher, M. W. (2001). The territorial integrity norm: International boundaries and the use of force. *International Organization, 55*(2), 215–250.

Appendix A. *Sources of Data*

1. S: Number of sovereign nation-states. Sovereign states as identified by Butcher and Griffiths (2020). Only countries that are independent today are shown. Current country borders are used across time. https://archive.ourworldindata.org/20251208-134119/grapher/sovereign-state-butcher-griffiths.html.
2. Y: total gross domestic product (output) of the world economy in billions of 2021 in international US dollars https://ourworldindata.org/grapher/global-gdp-over-the-long-run
3. N: total world population. Country data is always based on current geographical borders. https://ourworldindata.org/grapher/population?utm_source=chatgpt.com
4. YN: per capita GDP=Y/N in thousands of 2021 international US dollars.
5. TO: Trade Openness: This index is defined as the sum of exports and imports, divided by gross domestic product (GDP). https://ourworldindata.org/search?q=globalization+over+5+centuries&resultType=all
6. E: Share of people aged 15 or older who received some formal primary, secondary, or tertiary education. https://ourworldindata.org/global-education
7. L: Share of adults who can read and write a simple statement about their everyday life. https://ourworldindata.org/grapher/literate-and-illiterate-world-population.
8. HR: Data by V-Dem. Expert estimates of the extent to which people are free from government torture, political killings, and forced labor, have property rights, and enjoy the freedoms of movement, religion, expression, and association. The index ranges from 0 to 1 (most rights). https://ourworldindata.org/human-rights.
9. DE: People living in autocracies (percentage of total population). https://ourworldindata.org/grapher/people-living-in-democracies-autocracies
10. PCI: a principal component index composed of Y/N, L and HR (see Appendix C)

Appendix B. *Raw Data*

Year	S	Y	N	YN	TO	E	L	HR	DE	PCI
1820	36	1628	1.090	1494	9.5	17.19	12.05	0.3329	0.85	-2.46
1830	46	1885	1.162	1622	10.0	18.53	13.39	0.3520	0.85	-2.152
1840	51	2014	1.220	1651	10.4	19.87	14.72	0.3648	0.86	-1.955
1850	53	2143	1.276	1680	10.9	21.22	16.06	0.3949	0.82	-1.65
1860	54	2431	1.298	1873	11.3	22.56	17.40	0.4158	0.84	-1.355
1870	53	2720	1.340	2029	11.8	23.9	18.74	0.4438	0.76	-1.052
1880	55	3787	1.415	2677	16.0	26.7	19.63	0.4469	0.75	-0.783
1890	54	4320	1.521	2841	17.0	29.6	20.52	0.4504	0.72	-0.676
1900	49	4854	1.627	2983	17.2	33.2	21.40	0.4078	0.76	-0.868
1910	53	5769	1.767	3264	19.7	35.6	26.45	0.4134	0.76	-0.57
1920	60	6684	1.895	3527	16.3	38.6	31.62	0.4169	0.47	-0.326
1930	63	8639	2.072	4170	12.4	40.9	32.53	0.4083	0.75	-0.23
1940	54	10594	2.292	4622	11.0	45.9	41.88	0.3584	0.80	-0.272
1950	77	11723	2.493	4702	19.9	54.34	35.96	0.4265	0.70	0.069
1960	103	18478	3.015	6128	22.4	60.52	41.62	0.4595	0.48	0.603
1970	132	30398	3.695	8228	25.7	67.12	54.73	0.4443	0.54	0.985
1980	157	44308	4.448	9962	40.0	72.8	67.85	0.4652	0.43	1.449

1990	166	59618	5.328	11190	38.1	77.57	74.91	0.5760	0.37	2.202
2000	191	78650	6.172	12744	47.0	81.13	80.98	0.6731	0.29	2.79
2010	194	112303	7.022	15994	56.3	84.69	84.23	0.6995	0.23	3.1
2020	195	147693	7.887	18726	51.7	86.94	86.75	0.6748	0.23	3.15

Note: The highlighted cells have been used using average between available data.

Appendix C. *Construction of the Principal Component Index (PCI)*

This appendix describes the construction of the Principal Component Index (PCI), which summarizes long-run global human development using three variables: per-capita GDP (YN_t), adult literacy (L_t), and the human rights index (HR_t).

Let:

$$X_t = (YN_t, L_t, HR_t)'$$

denote the vector of raw variables observed at decade $t = 1820, \ldots, 2020$. Because the variables differ in scale and units, each series is standardized prior to analysis:

$$Z_{it} = \frac{X_{it} - \bar{X}_i}{\sigma_i}$$

where $\bar{X}_i$ and σ_i are the sample mean and standard deviation of variable i. Let:

$$Z_t = (Z_{\mathrm{YN,t}}, Z_{L,t}, Z_{HR,t})'$$

be the standardized vector.

Principal component analysis (PCA) is then applied to the 21×3 matrix of standardized observations. The sample covariance matrix is used to extract the principal components.

$$\Sigma = \frac{1}{T-1}\sum_{t=1}^{T}(Z_t - \bar{Z})(Z_t - \bar{Z})'$$

and is decomposed by solving the eigenvalue problem:

$$\Sigma v_k = \lambda_k v_k.$$

The first eigenvalue λ_1 explains 94.7% of the total variance, indicating a dominant single factor.

The associated normalized eigenvector is:

$$v_1 = (0.586, 0.571, 0.576)'$$

The PCI for each decade is computed as the linear combination:

$$PCI_t = v_1' Z_t = 0.586\, Z_{YN,t} + 0.571\, Z_{L,t} + 0.576\, Z_{HR,t},$$

All three loadings are positive and nearly equal, consistent with interpreting the PCI as a composite measure of economic modernization, human capital accumulation, and institutional liberalization. Breakpoint-augmented Dickey–Fuller tests confirm that the resulting PCI is stationary, making it suitable for inclusion in level-based regressions without inducing spurious correlation.