The Statistical Casualties of Autocratization: Mapping Data Deterioration and Its Impact on SDG Performance

Sevastian Sanchez
Columbia University, Quantitative Methods in the Social Sciences
GR 5999 Master's Thesis
Spring 2025
Dr. Kenneth J. Grossberger

Author Note

[Include any grant/funding information and a complete correspondence address.]

TABLE OF CONTENTS [NOT UPDATED]

ABSTRACT

As democratic institutions weaken globally, what happens to a country's ability to collect and use quality data, and how does this affect the delivery of essential services like healthcare, education, and clean water that billions depend on? This thesis investigates how democratic backsliding affects statistical capacity—a country's ability to produce reliable data—and consequently, progress toward achieving the UN Sustainable Development Goals (SDGs). While research has established correlations between democracy and statistical performance (70%) and between democracy and SDG achievement ($R=0.648$, $p<0.01$), limited attention has been paid to how statistical systems respond during democratic regression (Di Gennaro, 2024; Ștefanachi et al., 2022). Notably, non-democratic countries show little correlation between democratic index scores and sustainable development outcomes ($R=0.04$, $p<0.05$), highlighting regime type's importance in development processes. Using longitudinal analysis with first-difference models and lagged variables, this study examines patterns of data deterioration in countries experiencing democratic backsliding, identifying which types of data diminish first and how this sequential loss impacts development outcomes. The research draws on three datasets: the World Bank's Statistical Performance Indicators measuring statistical capacity across five pillars for 181 countries (2016-2022); the UN SDG Index tracking performance on 17 goals (2000-2023); and the V-Dem Dataset with Episodes of Regime Transformation capturing regime transitions. By focusing on Data Use and Data Quality dimensions, this study addresses a critical gap in understanding how governments mobilize data during regime transitions. The findings will contribute to both academic understanding of the democracy-development nexus and practical policy interventions aimed at preserving statistical independence during political instability. This research is motivated by the recognition that "statistics have value only if they are used widely and frequently" (Di Gennaro, 2024) and that "the availability of standards does not equal the availability of data" (Kitzmueller et al., n.d.). [PLACEHOLDER ABSTRACT]

Keywords: [Tap here to add keywords.]

INTRODUCTION

In a speech given to kick off San Francisco's Climate Week, former Vice President Al Gore quoted German philosopher, Jürgen Habermas, who famously stated that, at the onset, Nazis "attacked the very heart of the distinction between true and false."
Ultimately, there is a universal truth from that statement.

- Authoritarian regimes don't exactly rise out government transparency and intellectual honesty.
- There is a reason why statistical offices are often perceives as 'enemies of the state.' When we control information, we control people.
- A key aspect of democracy is contestation. Contestation is enshrined in civil liberties allotted to citizens, granting fundamental rights not limited to freedom of press, religion, protest, etc. Statistical capacity, although not too often acknowledged as a primal asset of democracy, is arguably one of the most important drivers of contestation.
- It is the very reason why statistical offices are often perceived as enemies of the state; why autocrats censor via book burnings, arrest educators, and defund institutions that check them (e.g., …)
- Censoring information via dismantling and/or manipulating national data-statistical systems and infrastructure is as strategic as it is repressive. National statistics, when uncorrupted, is a powerful political institution that checks the government by providing means of inspection. Policymakers, civil society/organizations and the general population alike maintain the unequivocal right to dissent and form evidence-based decisions to act against the existing government. A system of checks and balances enables the pursuit of contestation.
- This study will evaluate the extent that good statistical/data systems are agents of democratic governance.

We often like to point to many different factors when describing development: GDP per capita, Gini coefficients, population, etc. However, we also tend to overlook the large role that the statistical/reporting system plays in facilitating development on its own, by capturing these pivotal agents, used to inform policy. In other words, we don't measure the measurer. Consequently, statistical offices are typically conceived as mere facilitators rather than essential drivers of impact. International organizations like the World Bank, UNDP and IMF gather country data, cast judgments on their reliability/adequacy (Source), and then officially disseminate data to dashboards and reports. These organizations use statistical indices, including but not limited to the SCI and, more recently, the SPI, to capture things like quality, availability and impartiality of statistical data. Like many other institutions, these indices reflect a broader pattern: democracies tend to exhibit stronger statistical systems, characterized by higher data quality, transparency, and accessibility. Research has shown a consistent relationship between democracy and statistical performance (e.g., Di Gennaro, 2024; Hollyer et al., 2010). This raises important questions about what happens to these systems when democratic institutions erode. Does democratic backsliding lead to a measurable decline in statistical capacity, and how might this affect broader development outcomes?

In order to find out, this examination is divided into three components. The first component operationalizes the established World Bank measures of statistical capacity, specifically *Statistical Performance Indicators (SPI)* and the *Statistical Capacity Indicators (SCI),* and the United Nations Sustainable Development Goals (SDGs)–Index to better understand how factors, not limited to data quality and infrastructure use, are linked to countries' progress on achieving different aspects of international development. This part aims to answer the following question: *To what extent is the strength of a country's statistical/tracking systems tied to its sustainable development progress?* This project assesses 'development' / 'development status' from the lens of the UN SDG framework because 'development' has a multitude of dimensions beyond the scope of prominent economic indicators like GDP per capita and life expectancy.[1]

The second component complements the first by assessing how statistical capacity directly affects the UN SDGs relationship over time, as cause and/or a mediating factor between countries' democracy levels and SDG progress. Little research has been done to link these three variables together, especially within the contexts on democratic backsliding.

- [expand on] Incorporates panel data techniques such as Fixed Effects and First Difference models.

The third will consist of specific case studies of countries that have experienced notable changes/transformations to preexisting governance and political structures that would otherwise withstand power shifts. [expand on].

[1] See Appendix A for additional information on the UN SDGs.

HYPOTHESIS & MOTIVATION

Provided existing literature and recent turn of events in the United States, I suspect there is a link between democratic governance and statistical capacity, which could in turn erode development capabilities of countries (e.g., diminishing health/medical data stalls research and development on critical treatments for impending illnesses or health crises). If democracy and statistical capacity are positively correlated to such a strong degree (Di Genaro, 2024), and if statistical capacity is a significant driver of development, then it is plausible to speculate that regime transition (autocratization / democratization) would indirectly effect development if democracy level impacts development statistics.

Divided into Four Hypotheses

The goal of this analysis is to assess the linkage between regime transition and sustainable development and to quantify the extent that statistical performance plays a role in suspected shifting of sustainable development outcomes. Accordingly, this study will be guided by the following hypotheses.

1. *Hypothesis 1:* Statistical capacity is tied to countries' SDG performance.
2. *Hypothesis 2:* Democratic backsliding negatively impacts statistical performance.
3. *Hypothesis 3:* A subsequent decline in statistical performance leads to slower or regressed progress on SDG outcomes.
4. *Hypothesis 4:* The relationship between regime transition (specifically democratic backslide) and SDG status is mediated by statistical capacity.

Why Does This Matter for Development?

The goal of this work is to create a way to empirically measure how changes in data systems (improved/worsened) impact overall development. This research could refine our understanding of how diminished data quality, from backsliding over time, can adversely affect people's lives: access to quality education, secure healthcare, academic freedom, etc. Additionally, it could provide insights about livelihood during democratic backsliding. For instance, losing disaggregated data representing minority groups can perpetuate social inequalities as regimes transition to fascism/authoritarianism. Additionally, concerns about data legitimacy can rise as citizens begin to question the trustworthiness of official statistics reported by an authoritative head of state.

Breaking down the rationale:

1. Empirically assessing the extent to which quality data/reporting fosters development progress: How much does investment in statistical capacity positively impact development? By uncovering the net impact of statistical capacity on development outcomes, we might better estimate which goals are more likely to benefit (on average) after an additional unit increase in statistical capacity. Understanding how

data/information acts during democratic backsliding could help predict policy and development outcomes.

2. Insight into what the quality of data can tell us about democratic backsliding: How does data quality reflect what is happening during regime change? This research aims to measure how changes in data systems impact overall development. It could help refine our understanding of how diminished data quality affects people's lives and provide insights about livelihood during democratic backsliding.

In summary, this research aims to quantify the impact of statistical performance on development outcomes and provide insights into the nature of democratic backsliding and its effects on essential services. By empirically assessing the impact of statistical capacity on development and the dynamics of data deterioration over regime change, governments and NGOs can be better equipped to sustain statistical independence and integrity over periods of instability and political turmoil.

EXISTING LITERATURE [EDITING ON SEPERATE DOC]

DATA COLLECTION & VARIABLES

To examine the relationship between regime change, statistical capacity and SDG performance (policy performance), this study requires measurable indicators that can best represent these variables and suit the panel-like structure necessitated by comparative analysis of countries over many years.

SDG Index, Sustainable Development Report (2024)

The UN Sustainable Development Goals are complemented by the *SDG Index*, which captures overall performance on 17 global policy benchmarks across 167 countries. As both global goals and measurable indices, the Sustainable Development Report's SDG index, includes both overall scores and individual scores for each of the 17 goals. In total, the 17 goals comprise of 169 targets, and 244 indicators (known to date) to achieve 2030 Agenda.[2] This study specifically focuses the Development Report's overall composite index score, comprising of 98 global indicators, while building a replicable procedure for future research on disaggregated SDG statistics (Appendix B.1). Although the SDGs were officially adopted in 2015, the SDG Index is backdated to the year 2000; these estimates lay the groundwork to compare statistical capacity measures that would otherwise vary significantly in sample size. That being said, it is important to consider the gaps in SDG data, which can impede estimates if not properly accounted for. For instance, SDG 14 ('Life Below Water') can be attributed to a general lack of water access in land-locked countries (e.g., Switzerland does not produce SDG 14 estimates), in turn, limited legal provisions, SDG 10 may have additional motivations such as a lack of statistical infrastructure/systems, national politics, and/or policy relevance. Luckily, the SDG index accounts for missing data by only calculating composite scores of countries that report more than 80% of testable indicators.[3] Country-years are estimates calculated at the __start/end?__of every consecutive year.

Statistical Performance Indicators (SPI) & Statistical Capacity Indicators (SCI)

The World Bank Group's Statistical Performance Indicators (SPI) assesses the statistical efficiency of 181 countries from 2016 to 2022.[4] This dataset includes six indices, consisting of one overall composite score, the Statistical Performance Index, and five individual pillar indices: Data Use, Data Services, Data Products, Data Sources, and Data Infrastructure. Together, these pillars are constituted by a total of 51 indicators. It is important to note, however, of the presence

[2] Refer to Appendix A for more information on the 2030 Agenda; See Appendix B.1 and Sustainable Development Report (2024) for more details on the SDG Index framework; data and methodology also accessible via their website: https://dashboards.sdgindex.org/downloads.

[3] According to the UN _[find specific source]__, "While all member states have a country profile, only those with less than 20% missing data are ranked in the SDG Index, resulting in 167 countries being ranked in the 2024 report."

[4] SPI data is extracted from the World Bank's Official GitHub Repository "SPI" and can be accessed here: https://github.com/worldbank/SPI/tree/master.

of data gaps specifically among records data starting from the year 2016. Conversely, the Statistical Capacity Index (SCI), also a World Bank development indicator, measures the statistical capabilities of 146 countries over time from 2004 to 2020.[5] Unlike the SPI, the SCI focuses solely on low-income and middle-income developing nations, excluding high-income countries. The SCI is based on 25 indicators, which is fewer than the 51 indicators considered by the SPI. To avoid selection bias, which would limiting the sample to all but what the World Bank Group classifies as 'developed countries,' this study leverages the Statistical Performance Index (SPI) as the main proxy / explanatory variable of statistical capacity.[6]

Economist Intelligence Unit, Democracy Index (EIU DI)

To confirm Di Gennaro's findings that there is a correlation between regime type and statistical/data capacity and uncover if/how changes in regime type affect statistical capacity, I start by extracting longitudinal data from the Economist Intelligence Unit's Democracy Index (DI).[7] The Democracy Index is based on 60 indicators, divided into five categories: *electoral process and pluralism, civil liberties, functioning of government, political participation* and *political culture.* Countries' overall composite scores are the average of these five individual scores, similar methodology to that of the Statistical Performance Indicators (SPI). Countries are rated on a continuous 0–10-point scale, which provides a robust and comprehendible starting point from where we can begin to compare scores with predictor variables (i.e., statistical capacity measures and controls). For this analysis, data is extracted from 15X countries in the years 2004-2023 to best match corresponding statistical capacity indicators and SDG composite scores. Ultimately, the index serves as a basis for future exploration into the topic of democratic backsliding and statistical erosion. Note that EIU Democracy Index is acknowledged as a starting point and is most certainly not the only measure of democracy. Still, to reiterate, EIU DI includes numerous indicators (e.g., accountability, civic participation, …) that collectively make up an overall score; accordingly, these indicators will not be specified in the models although their involvement is assumed.[8]

General V-Dem & Events of Regive Transition (ERT), Varieties of Democracy (v14)

These two datasets, produced by Varieties of Democracy (V-Dem), together contain statistical measures on countries' level of democracy, conceptualized through 5 different variations, or facets, of democracy: liberal democracy, electoral democracy, participatory democracy, egalitarian democracy and deliberative democracy.[9] The Variety of Democracies' Environment of Regime Transition (ERT) dataset specifically contains regime transition variables used to, 1) isolate countries which undergone or are undergoing democratic backsliding between the years 2004-2023; and 2) compare transitioning regimes, specifically in identifying

[5] SCI data is extracted from World Bank Groups "Data on Statistical Capacity" webpage, accessible here: https://datacatalog.worldbank.org/search/dataset/0037854/data-on-statistical-capacity.

[6] Refer to Appendix B.2 & B.3 for a greater breakdown of the Statistical Performance Index (SPI) and Statistical Capacity Indicators (SCI).

[7] Economist Intelligence Unit (2006-2024); data was accessed through Our World in Data, third-party source, found here: https://ourworldindata.org/grapher/democracy-index-eiu; EIU Democracy Index reports can be accessed here:

[8] See Appendix B.4 for more details on EIU Democracy Index indicators.

[9] Accessed through: V-Dem Website, https://www.v-dem.net/data/the-v-dem-dataset/country-year-v-dem-fullothers-v14/

similar and/or unique patterns observed over time – operationalized across different statistical longitudinal models/tests.

Although literature surrounding regime transition tends to focus on elections, just/unjust transitions of power and political/economic preconditions (SOURCE), regime change, especially in the direction of autocratization, can be unexpected and invoked by a plethora of events requiring research beyond the scope of this analysis. As such, this paper refrains from casting any kinds of predictions on democratic backsliding and will focus primarily on defined instances of backsliding that occurred within the 2015–2023-year range. This study applies the following pre-established regime change indicators developed by V-Dem scholars found in the ERT dataset:

[Integrate the following]

- Aut_ep:
- Dem_ep:
- Regch_event: A categorical indicator, V-Dem's RoW regime change event (regch_event) identifies occurrences of consistent decrease/increase in countries' democracy levels that result in a complete change in regime status. Based Regimes of the World (RoW) methodology, the regime change event variable assesses the direction of transition (towards democracy vs. autocracy) and the specific country-year when transition first occurs for such given country. Unlike regch_genuine, regch_event does not necessitate a founding autocratic/democratic election to recognize a change in regime status. For the contexts of this study, founding elections aren't necessary to understand how statistical capacity reacts to *incremental* changes in democracy levels.

The incrementalist assumption…

Descriptive Summary Statistics of Key Variables

Furthermore, the number of sample observations changes between the three components due to their individual design structures. Applying pooled OLS techniques allows for all country data, regardless yearly of gaps, to be analyzed simultaneously. That said, consecutive country-year estimates in variables (i.e., democracy score, SDG index, statistical capacity and controls) within at least 7-10 years is the criterion by which countries are suitable for Component-Two's panel data analysis.

Table 1

Sample-Specific Statistics of Indices

	SDG Score	SPI Score	SCI Score	Democracy Index
Years	2004 - 2023 (20)	2016 - 2023 (8)	2004 - 2020 (17)	2006 - 2024 (16)
N countries	167	165	123	160
N country-years	3360	1300	2072	2560
NA	0	2060	1288	800

Note. EIU's Democracy Index does not produce estimates in the years 2007 and 2009. This study does not attempt to fill missing years; pooled OLS estimates are calculated without imputation techniques to mitigate bias.

Table 2

Summary Statistics by Indicator

Indices (composite)	Min	1st Qu.	Median	Mean	3rd Qu.	Max	N (countries)	NA's
SDG (0-100)	35.65	55.74	65.53	64.43	73.05	86.42	168	0
SPI (0-100)	11.77	54.82	67.09	66.71	81.48	95.26	187	2060
SCI (0-100)	16.67	57.78	70.00	67.96	80.00	98.89	146	1288
DI (0-10)	0.26	3.52	5.78	5.51	7.280	9.93	168	800

Note. All indices seem relatively standard, with relatively little deviation between mean and median estimates. [discuss NA’s]

RESEARCH DESIGN AND METHODOLOGY

To reiterate, democracy, as an ideal-type, is a regime type characterized by its strong statistical reporting, high-quality disaggregated statistics and data transparency.[10] While democracy does not necessarily translate to “development,” a statistically significant relationship between statistical capacity and democracy exists (Di Genero, 2024), which raises interesting questions about the role of data in democratic consolidation. If there is a statistical link between democracy and reliable official statistics, will official statistics diminish as countries experience democratic backsliding? Will statistical capacity and, by association, the UN SDG score trends mutually fluctuate downwards during the process? Answering these questions requires deeper inspection of statistical capacity measures, specifically to identify which is most suitable for capturing the complex and multi-faceted nature of the UN Sustainable Development Goals.

Guiding Research Questions

1. Is statistical capacity a strong predictor of SDG status? What is the best measure of statistical capacity?
2. Do shifts in democracy levels result in shifts in statistical capacity and SDG status?
3. How does this process look over time, comparing regressing regimes? Are there similar patterns?

This study is designed to bridge the elements of policymaking, national data/statistics and regime change together and solidify a foundation for further analysis on how they influence each other over time. Accordingly, this study refrains from making any causal claims as to how $x \rightarrow y$ before providing sufficient empirical justification. Thus, Component 1 is particularly designated to examining if and how statistical capacity (x) affects SDG performance (y); this study interrogates two primal measures of statistical capacity: Statistical Performance Indicators (SPI) and Statistical Capacity Indicators (SCI). Their composite index scores (i.e., Statistical

[10] source

Performance *Index* and Statistical Capacity *Index*), are aggregates of several outlined components that collectively constitute each measure. Both are compared systematically via cross-sectional pooled regression techniques before transitioning to longitudinal panel analysis.

Component 2 of this study scrutinizes the impact of statistical capacity on overall SDG performance, with a particular focus on variable changes and how dependent variables (i.e., statistical capacity and SDG performance) react to shifts in democracy levels of countries. This longitudinal-panel inspection will aim to uncover any significant shifts that occur in the correlation between SDG status and statistical capacity and establish a theoretical basis before delving into specific case studies of countries that have and/or are actively experiencing democratic backsliding (2015-2023). We then discern similar patterns in statistical capacity and developmental outcomes of the following countries: ___ [List choice of 4 countries here].

To ensure the robustness of the findings, various statistical checks will be employed, not limited to correlation matrices, assessments of multicollinearity using Variance Inflation Factor (VIF) analysis, and custard-grouped standard errors for heteroskedasticity. Through these, and other, comprehensive analyses, the study seeks to provide insights into the intricate relationship between statistical capacity and SDG performance, both generally and amid regime transitions.

COMPONENT 1: REGRESSING MEASURES OF STATISTICAL CAPACITY WITH SDG PERFORMANCE

What is the best statistical capacity measure for predicting SDG Progress? What is the overall relationship between Statistical Capacity and SDG progress?

Identifying Adequate Measure(s) of Statistical Capacity

H0: No relationship
H1: Statistical capacity is tied to countries' SDG performance [two-tailed]

The objective of this research design and analysis is multifaceted. This first component aims to thoroughly evaluate two measures of statistical capacity to identify the most appropriate/versatile predictor of Sustainable Development Goal (SDG) performance. In addition to the Statistical Performance Index (SPI), this paper considers the Statistical Capacity Indicators (SCI). Simultaneously, this section reveals a rough approximation of the overall impact of statistical capacity factors (i.e., data gaps, data quality, data transparency, underreporting/tracking, data infrastructural shortcomings) on countries' ability to address pressing issues, such as food security (SDG 2), quality healthcare/services (SDG 3), and electricity and broadband (SDG 7). Of course, this is in no way to suggest that all of the 17 scores will move in the same direction. In fact, a crucial goal of this analysis is to arrive at a point where disaggregating the main dependent variable (*Y = SDG composite score*) into 17 sub scores, which are predetermined using SDG Index methodology (Appendix X).

As outlined below, this segment comprises of a preliminary analysis and three inferential statistical analyses, and corresponding robustness checks, to assess whether *statistical capacity impact SDG performance (H1).*

Design Overview [To be Updated]

1) **Preliminary Analysis**
2) **Test 1:** Multiple POLS: SDGs ~ Statistical Capacity Measure + Controls
   a) VIF – checks for multicollinearity
   b) Visual inspection
3) **Test 2:** Multiple POLS: Non-linear Quadratic Terms
   a) Evidence of Misspecification / non-linearity
   b) Model Diagnostics
4) **Test 3:** Multiple POLS: SDGs ~ SPI + Moderator + SPI(Moderator) + Controls
   a) Evidence of Interactions
   b) Model Diagnostics
   c) [Subgroup Analysis] Multiple POLS: SDGs ~ SPI + Controls
      i) High Income Countries
      ii) Upper Middle-Income Countries

iii) Lower Middle-Income Countries
iv) Low Income Countries

Causal Chain: [CREATE IN R]
Democracy/Regime Type → ***Statistical Capacity*** → ***SDG Status***

Preliminary Analysis: Correlation & Naive OLS Models

To aid the following theoretically justified strategy, some preliminary examination on the statistical capacity and SDG performance relationship was conducted (Appendix X). In summation, it confirmed the relatively strong pearson correlation (r) between the Statistical Performance Index (SPI) and the SDG Index (r = 0.7849); in a bivariate regression, the model accounts for about 61.6% of the variance between SPI and SDG performance ($R^2 = 0.6160$). As for the Statistical Capacity Index (SCI), based on the data (Appendix X), the correlation with SDG performance is roughly 0.647, which explains about 41.8% percent of the variation between SCI and SDG when squared ($R^2 = 0.41796$). The impact of SCI on SDG and SPI on SDG are statistically significant, in all Naïve Ordinary Least Squares models (Appendix X).

TEST 1: Multiple Pooled Ordinary Least Squares + Robustness Checks

All variables of statistical capacity (SPI & SCI) will be compared on a base pooled OLS regression model structure. Pooled OLS recognizes the panel-like structure allowing to index by specific country and year (country-year). Regular OLS, assumes independence of observations which is not suitable given the repeated waves of country-year units over the course of multiple consecutive years (discussed under Anticipating Multicollinearity and Omitted Variable Bias sections).

Model Comparisons. The following models evaluate the strengths of different possible predictors (X-variables) of SDG status (Figure X). Models 1 and 2 compare relationships of both Statistical Performance Index (SPI) and Statistical Capacity Indicators (SCI) measures with the UN Sustainable Development Goals (SDG), holding all else constant. Like the preliminary design (Appendix X), model 3 combines predictors of statistical capacity (i.e., SPI & SCI) into a single multiple regression model to explore its potential in sustaining a stronger model.

Control Variables. For consistency, models 1, 2 and 3 adopt the same controls: EIU's Democracy Index (*di_score*), Log of GDP Per Capita (*log_gdppc*), and the year (*factor(year)*). Because predictor variables SPI, SCI and DI are aggregate scores of individually combined indicators, this analysis does not necessitate separate indicator control variables. Including a variable that is already used to build an aggregate index is like putting eggs into a cake recipe twice—you're double-counting the same ingredient. To avoid this, I exclude individual indicators that are already part of the composite measure.[11]

- Controls: Are they experiencing regime change (Y/N)

[11] See Appendix B for a full list of indicators that collectively constitute aggregate index calculations of SPI, SCI and DI.

Time Period. Although SDG data dates to the year 2000, this country-year analysis starts at 2004 to match the earliest measure of statistical capacity, specifically the Statistical Capacity Index (SCI) which starts in 2004.

Test 1 not only serves to identify the most impactful indices of statistical capacity, but it also helps to understand how well a country's statistical performance/capacity correlates with its progress towards the UN SDGs overall.

Hypothesis 1 is broken down into several sub-hypotheses, specifically:

- ***H0:*** Null, no relationship(s) found
- ***H1a:*** There is a positive, strong and statistically significant relationship between the 17 UN SDGs and Statistical Performance Index (SPI) [one-tailed]
- ***H1b:*** There is a positive, strong and statistically significant relationship between the 17 UN SDGs and Statistical Capacity Index (SCI) [one-tailed]
- ***H1c:*** Statistical Performance Index (SPI) is a stronger predictor of SDG Status compared to Statistical Capacity Index (SCI) [one-tailed]

Model 1: ols_spi

$$SDG\ Performance_{Overall} = a + \beta_1 SPI_{Overall} + \beta_2 Democracy\ Index_{Overall} + \beta_3 Log\left(GDP_{Per\ Capita}\right) + \beta_4 Year + \epsilon$$

Model 2: ols_sci

$$SDG\ Performance_{Overall} = a + \beta_1 SCI_{Overall} + \beta_2 Democracy\ Index_{Overall} + \beta_3 Log\left(GDP_{Per\ Capita}\right) + \beta_4 Year + \epsilon$$

Model 3: ols_multiple

$$SDG\ Performance_{Overall} = a + \beta_1 SPI_{Overall} + \beta_2 SCI_{Overall} + \beta_3 Democracy\ Index_{Overall} + \beta_4 Log\left(GDP_{Per\ Capita}\right) + \beta_5 Year + \epsilon$$

Model Misspecification and Robustness Checks. This study anticipates different forms of model misspecification.[12] To test for multicollinearity, we employ Variance Inflation Factor (VIF) assessments of predictors in all three models.[13] Highly correlated variables, although seemingly relavent to the underlying theory behind this study (e.g., countries' total population), are omitted for interpretation purposes.

Additionally, it is crucial to address both non-linearity (functional form misspecification) and omitted variable bias in the pooled OLS model, as both can distort the estimated relationships between statistical capacity and SDG progress. Models are specified with relevant

---

[12] As noted by Wooldridge, "[m]isspecification analysis is the process of determining likely biases that can arise from omitted variables, measurement error, simultaneity, and other kinds of model misspecification" (*Introductory Econometrics: A Modern Approach*, see glossary and Ch. 9).

[13] Variance Inflation Factor, or VIF, examines the variance of regression coefficients of a model, particularly the extent that it increases because of collinearity found among a model's independent variables. A VIF estimate above 5 is typically understood as evidence of moderate multicollinearity, while values above 10 is considered severe. For more details, see Wooldridge (2013), *Introductory Econometrics: A Modern Approach* (pp. 96–98).

controls to mitigate omitted variable bias, but alternative panel techniques are necessary to fully account for unobserved time-invariant factors.[14] To assess whether non-linear terms improve model fit, the Ramsey RESET test (Ramsey, 1969) is conducted at the start of Test 2.[15]

Having conducted Breusch–Pagan tests for heteroskedasticity (Breusch & Pagan, 1979), results indicate clear evidence of non-constant variance of residuals found in all three models (Appendix X). This is expected given the real-world panel structure of the data. To address it – and all cases of potential heteroskedasticity and within-country correlation – all models in first and second component apply robust clustered standard errors at the country level using the HC1 correction (White, 1980); this would sustain the assumption of independence between countries.

Hypothesis Testing & Model Diagnostics. Using robust standard errors clustered at the country level, this analysis test whether statistical capacity significantly predicts SDG performance. The t-test coefficients for the statistical capacity measures, t-statistic and p-values of predictors, determines the existence of evidence, whether strong, weak or non-existent, that statistical capacity actually matters for SDG performance.

Admittedly, Test 1 assesses fit based on surface-level analysis rather than a deeper examination into specific indicators or pillars and how they correlate with SDG composite and sub scores. We discern the 'best model' as one that has the lowest AIC/BIC scores and highest Adjusted R^2 at an aggregate level – the defining factors by which the most appropriate model is selected. While this surface-level approach a notable limitation of this study (refer to Discussion: Limitations and Opportunities), descriptive statistics of indices is weighted against said factors and ultimately informs how this study determines a primary explanatory measure of statistical capacity (Appendix X).

TEST 1: Results and Analysis

Table 3

Summarizing Results of Models 1, 2 and 3

	Dependent variable:		
	SDG Performance (composite)		
	ols_spi (Model 1)	ols_sci (Model 2)	ols_multiple (Model 3)
SPI Composite	0.287^{***}		0.124^{**}
	(0.038)		(0.057)
SCI Composite		0.238^{***}	0.139^{***}

[14] Component 2 of this study navigates alternative quantitative models such as *Fixed Effects* and *First Difference*. These techniques ensure time-invariant factors, including geographic region, colonial history, culture, etc., are accounted for.

[15] The Regression Equation Specification Error Test (RESET) test assesses whether non-linear arrangements of explanatory terms could possibly refine the model fit, essentially to better help explain the dependent variable. For more information on the Ramsey RESET test, see Ramsey (1969).

		(0.030)	(0.048)
DI Composite	0.206	-0.083	-0.392
	(0.285)	(0.268)	(0.276)
Log (GDP per Capita)	3.312***	5.508***	5.793***
	(0.475)	(0.474)	(0.463)
Year 2008		-1.250***	
		(0.236)	
Year 2010		-1.019***	
		(0.297)	
Year 2011		-1.505***	
		(0.361)	
Year 2012		-1.035***	
		(0.391)	
Year 2013		-0.956**	
		(0.418)	
Year 2014		-0.551	
		(0.430)	
Year 2015		0.264	
		(0.412)	
Year 2016		0.743*	
		(0.419)	
Year 2017	-0.419***	0.956**	-0.072
	(0.142)	(0.453)	(0.188)
Year 2018	-0.959***	1.486***	0.050
	(0.240)	(0.452)	(0.357)
Year 2019	-0.631**	1.888***	0.400
	(0.262)	(0.462)	(0.398)
Year 2020	-0.655*	2.983***	1.021
	(0.360)	(0.484)	(0.621)
Year 2021	-2.196***		
	(0.518)		
Year 2022	-2.110***		
	(0.493)		
Year 2023	-2.495***		
	(0.525)		

Intercept	18.969***	1.881	1.968
	(2.518)	(2.951)	(3.053)
Observations	1,236	1,515	567
R^2	0.778	0.744	0.760
Adjusted R^2	0.776	0.742	0.757
Note.			*p<0.1; **p<0.05; ***p<0.01

The results from Test 1 multiple pooled OLS models – with clustered robust standard errors – indicate that both SPI and SCI have a statistically significant relationship with SDG composite scores. The p-values for both SPI and SCI are less than 0.001, indicating strong evidence against the null hypothesis of no relationship for either model. Holding all else constant (log GDP per capita, democracy score and year), SPI and SCI exhibit positive moderate and statistically significant relationships with SDG status.

Upon comparing separate models, SPI has a greater impact on SDG status (0.28735) than SCI (0.237633). This suggests that a one-unit increase in SPI is associated with a larger improvement in SDG outcomes compared to a one-unit increase in SCI, holding all controls constant. Interestingly, the opposite holds true in a multiple regression model containing both SPI and SCI. SPI's impact on SDG status (0.124233) (net of SPI) is less than that of SCI's (0.138766) (net of SCI), holding all controls constant. When together, the coefficients represent the unique impact of each predictor variable (measures of statistical capacity) on SDG status, net of all other variables.

Model 1 (ols_spi) does not control for SCI and model 2 (ols_sci) does not control for SPI – this is okay. SPI is the predecessor of the SCI, sharing overlapping data; significant statistical correlation (multicollinearity) is expected. This is possibly what explains the reduction of both regression coefficients in model 3: 0.28735 to 0.124233 for SPI (56.77% decrease); and from 0.237633 to 0.138766 for SCI (41.60% decrease). This indicates that both variables capture much of the same underlying relationship with SDG performance. However, the fact that both SPI and SCI remain significant when included together (model 3), although SCI less so than SPI, with a high adjusted R^2 value (0.741525), which suggests they capture different dimensions of statistical capacity that independently contribute to SDG status.

Collinearity & VIF Results. The correlation between SCI and SPI is approximately 0.828. When placed within the same model, SCI inflated the standard error of SPI from 0.0131 to 0.0271. SCI had a similar reaction from the SPI with its standard error increasing from 0.0096 to 0.0241. Such multicollinearity is moderately reflected in the VIF test which accounts for all explanatory variables instead of only statistical capacity measures (SPI & SCI).

Table X

Variance Inflation Factors (VIF) – Model 3: ols_multiple

Term	VIF	Df	$GVIF^{1/(2*Df)}$
spi_comp	4.316894	1	2.077714
sci_overall	3.754613	1	1.937682
di_score	1.598314	1	1.264244
log_gdppc	1.485085	1	1.218641

factor(year)	1.302733	4	1.033611

Note. VIF scores < 5 signify low collinearity; scores ≥ 5 signify moderate collinearity; and variables with scores 10 ≥ exhibit severe collinearity. GVIF scores are only relevant for categorical variable (i.e., factor(year)) and serve the same purpose as VIF.

Overall, there reveals no severe multicollinearity (all VIF < 5). There is moderate correlation between statistical capacity measures (spi_comp and sci_overall) with SPI moderately inflated by a factor of 4.32 and SCI inflated by a factor of 3.75. This alone signals weak evidence of multicollinearity, making it acceptable to include both predictors in the same model as doing so would not severely impact model estimates with both factors less than 5.0. Even though all other variables show minimal multicollinearity concerns, there are significant limitations in either model, including sample size, and longitudinal suitability.

Considering Missing Data – Systematic, Non-Random Pattern. SPI has near complete country data coverage (165 out of 168 countries with an SDG score), but with a stubborn temporal limitation (2016-2023). On the other hand, SCI has longer temporal coverage (2004-2020) but lacks reporting on high-income countries focusing primarily on the developing world (123 out of 168 countries with an SDG score).

The Best Model = Most Suitable Statistical Capacity Measure. To determine which predictor (i.e., SPI or SCI) is stronger –also considering if they're stronger *together* (*ols_multiple*)– this investigation considers (a) descriptive statistics (i.e., number of observations); (b) adjusted R^2 estimates; and (c) AIC/BIC model scores.[16] These measures are compared across models that sustain the same controls, however, statistical checks are weighted against frequency of observations, given anticipated differences in the number of years and mis-alignment of articular start to end year periods.

Recall the following hypothesis:

- *H0: Null, evidence against SPI > SCI*
- *H1c: Statistical Performance Index (SPI) is a stronger predictor of SDG performance compared to Statistical Capacity Index (SCI)*

Table X

AIC/BIC Results & Adjusted R2 & of OLS Models

Model	AIC	BIC	Adj. R^2	N
(1) ols_spi	7382.060	7443.496	0.776	1,236
(2) ols_sci	9003.178	9093.672	0.742	1,515
(3) ols_multiple	3292.001	3335.405	0.757	567

Note. For reference, smaller AIC and BIC scores indicate a better fit. Accordingly, the following one-tailed hypothesis is being tested: H0: Null, *ols_sci* (SCI) and/or *ols_multiple* (together) > *ols_spi* (SPI); H1c: *ols_spi* (SPI) > *ols_sci* (SCI) & *ols_multiple* (together).

Model 3 (*ols_multiple*) sacrifices a significant portion of its sample size in order to include both statistical capacity measures (SPI & SCI). While model 3 appears to outperform the

[16] AIC/BIC ….

other two models, the lower AIC/BIC comes at the expense of a significantly smaller number of country-year observations, which may not make it the most suitable fit for longitudinal analysis (i.e., looking to Component 2). Model 1 (*ols_spi*) considers a greater number of countries compared to model 2 (*ols_sci*),[17] and sustains slightly higher explanatory power than model 2 and model 3 (Adj. R^2: 0. 776 > 0.742; 0. 776 > 0.757, respectively).

Model 3 has lower AIC/BIC scores than model 2 (AIC/BIC: 3292.001, 3335.405 < 9003.178, 9093.672). As mentioned before, however, model 2 has a much smaller sample size, which significantly impedes results. Models 1 and 2 have significantly more country-year data points (n=1236 and n=1515, respectively) for regression analysis compared to model 3 (n=567). With all else considered, this study employs the Statistical Performance Index (SPI) as the primary measure of statistical capacity.

TEST 2: Non-Linear Misspecification and Interactions

Following standard econometric practice, I first established the appropriate functional form for main effects using Ramsey RESET tests (1969) and residual diagnostics, then I test for theoretically motivated interaction effects that could influence the relationship behind statistical capacity and SDG performance.

Procedure

1) Ramsey RESET Test for Misspecification & Omitted Interactions (non-linearity)
2) Check residuals for non-linearity and heteroskedasticity:
   - Residuals vs Fitted Values plot (for each model)
   - Residuals vs Predictor plot (for each predictor)
3) Establish main effects with pooled model with robust SE
4) Establish non-linear (quadratic, interactions, logged, etc.) model with robust SE (all continuous are quadratic)
5) Reduce the model by dropping non-significant terms (if any)
6) Compare AIC/BIC & Adjusted R-Sq values for both pooled and reduced model
7) Repeat steps 1-6 for Test 3 – Interactions

Test 2a: Ramsey RESET & Residual Diagnostics for Non-Linearity of Predictors

The Ramsey RESET test results (Appendix X) indicate that the null hypothesis of no omitted variables or non-linear relationships can be rejected for all three models (SPI, SCI, and Multiple). The p-values are all less than 0.05, suggesting that there are indeed non-linear relationships or omitted interactions in the models. The functional form of the models may not be correctly specified, and further investigation into non-linear terms or interactions is warranted.

Ştefanachi's et.al. (2022) findings demonstrated non-linearity, which further validates the RESET test results suggesting that democracy and sustainable development, in particular, maintain a non-linear relationship. Accordingly, this study performs residual analysis of continuous variables of *Residual-Fitted* and *Residual-Predictor* relationships. Figure X below

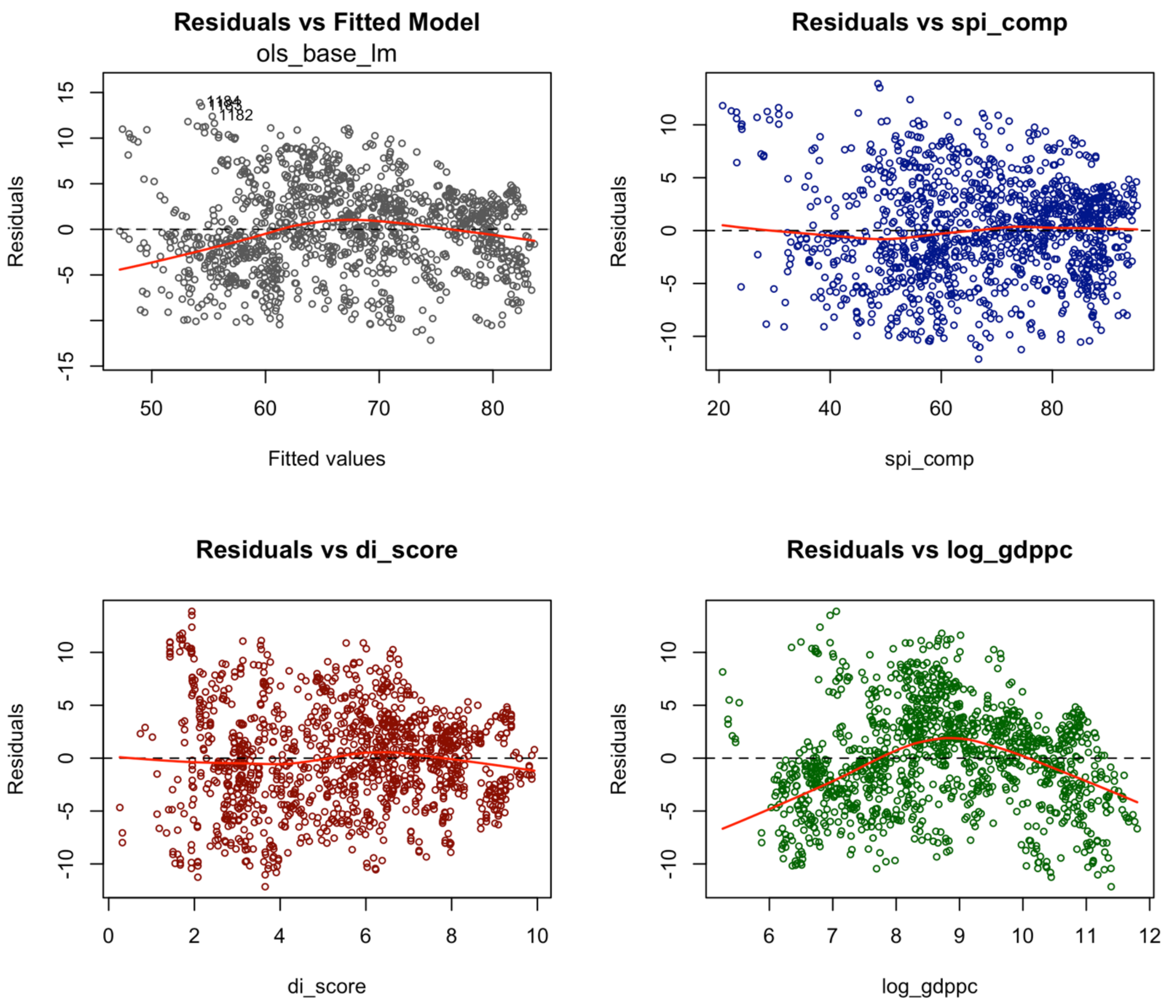
---

[17] Refer to Tables 1 and 2 in Data Collection & Variables for the breakdown of countries and country-year counts covered in each variable index.

displays four residual plots, the first (top-left) displaying the residual deviations across all continuous predictors. The next three isolate and visualize the non-linear effect of each predictor on SDG performance.

Figure X

Residual-Fitted and Residual-Predictor Plots of Main Predictors



After log-transformation, GDP Per-Capita displays an inverted U-shape, which is the most defined among all other continuous predictors. This signifies the non-linear nature of GDP Per-Capita, and the need to apply quadratic and/or other non-linear functional forms into the regression model to better proxy economic development. Empirically, this would mean including a quadratic term for log-transformed proxy for economic development:

$$SDG\ Performance_{Overall} = a + \beta_x \ldots + \beta_5 Log\left(GDP_{Per\ Capita}\right) + \beta_6 Log\left(GDP_{Per\ Capita}\right)^2$$

Test 2b: Quadratic Terms

Given the results of Test 1, and the evidence of misspecification found across all pooled OLS models (Appendix X), the following analyses sustain the Statistical Performance Index (SPI) as the prevailing statistical capacity measure. The corresponding model, depicted below (Fig. X), is utilized as a baseline to a more refined model capable of overcoming misspecification.

$$SDG\ Performance_{Overall} = a + \beta_1 SPI_{Overall} + \beta_2 Democracy\ Index_{Overall} + \beta_3 Log\left(GDP_{Per\ Capita}\right) + \beta_4 Year + \epsilon$$

Next, we formulate a quadratic model (*ols_spi_quad*) containing second-order quadratic terms for all continuous variables (equation/figure below). If the quadratic term of any continuous x-variable (i.e., SPI, DI, Log($GDP_{per\text{-}capita}$)) has a p-value ≥ 0.05, it suggests insufficient evidence for improving model fit. Consequently, the quadratic term would be dropped from the model, and the variable would retain linear specification.

$$SDG\ Performance_{Overall} = a + \beta_1 SPI_{Overall} + \beta_2 SPI^2_{Overall} + \beta_3 Democracy\ Index_{Overall} + \beta_4 Democracy\ Index^2_{Overall} + \beta_5 Log\left(GDP_{Per\ Capita}\right) + \beta_5 Log\left(GDP_{Per\ Capita}\right)^2 + \beta_4 Year + \epsilon$$

Comparing the linear-baseline and quadratic pooled OLS models, with robust clustered standard errors, the quadratic specification provides a significantly better fit for explaining SDG performance. In the linear model, statistical capacity (*spi_comp*) and log GDP per capita (*log_gdppc*) are both highly significant predictors, while the democracy index (*di_score*) is not significant. However, after adding quadratic terms for all three continuous predictors, we uncovered evidence of a non-linear underlying relational structure. All continuous explanatory variables (*SPI, DI and Log(GDP Per-Cap)*) are statistically significant (as quadratic variables), suggesting a non-linear relationship exists between predictors and SDG performance. The quadratic model shows a meaningful improvement in explanatory power, with the adjusted R-squared improving from 0.7757 to 0.8202 (4.45 percentage point increase in adjusted R^2).

To reiterate, the linear term for democracy index, previously non-significant, becomes significant when its quadratic is included, suggesting a non-linear relationship. The linear term for statistical capacity loses significance, but the quadratic remains, which—according to the Nelder's marginality principle (1977)—means both should be retained.[18] Overall, these results indicate that all three predictors exhibit important non-linear effects on SDG performance, which, though this logic, upholds the quadratic form as the superior model.

Table X

Pooled OLS Models with Quadratic Terms
[FUTURE-EDIT TO REFLECT CENTERED MEAN TERMS]

	Dependent Variable: SDG Performance (composite)

[18] All interaction and quadratic terms across all models include their respective lower-order variable terms due to the marginality principle; see Nelder's *A Reformulation of Linear Models* (1977).

	Linear-Baseline	1 Quadratic	2 Quadratic (A)	2 Quadratic (B)	3 Quadratic
spi_comp	0.287***	0.284***	-0.101	0.292	0.0005
	(0.038)	(0.036)	(0.165)	(0.033)***	(0.150)
spi_comp^2			0.003**		0.002**
			(0.001)		(0.001)
di_score	0.206	0.313	0.300	-2.363	-1.946**
	(0.285)	(0.259)	(0.251)	(0.957)**	(0.911)
di_score^2				0.266	0.223***
				(0.089)***	(0.084)
log_gdppc	3.312***	17.357***	18.519***	22.018	22.149***
	(0.475)	(3.375)	(3.446)	(3.698)***	(3.669)
log_gdppc^2		-0.808***	-0.892***	-1.104	-1.120***
		(0.189)	(0.194)	(0.210)***	(0.208)
Intercept	18.969***	-40.805***	-33.370**	-53.471	-45.840***
	(2.518)	(14.684)	(15.018)	(15.233)***	(15.428)
Observations	1,236	1,236	1,236	1,236	1,236
R^2	0.778	0.809	0.816	0.818	0.822
Adjusted R^2	0.776	0.807	0.814	0.817	0.820
AIC BIC	7446.06, 7671.324	7266.005, 7511.748	7228.738, 7494.959	7211.045, 7477.266	7193.846, 7480.546

Note. All models apply heteroskedasticity-robust clustered errors. *p<0.1; **p<0.05;***p<0.01

Although these results show meaningful improvements in the quadratic model, it is important to note that VIF estimates for the quadratic terms are higher than those for the linear terms, signaling multicollinearity issues (Table X). The severity of VIF results (all > 10) suggests that quadratic terms are inflating the variance of the coefficient estimates. Assessing which variable(s) is causing the most non-linearity is essential to determine if the quadratic terms are necessary. To address this, we *mean-center* continuous predictors – their linear form and quadratic – by subtracting the variable mean ($\bar{x}$) from corresponding observed values. This process is written as $x_i - \bar{x}$, where x_i denotes an observed value of a given predictor.

Table X

Variance Inflation Factors (VIF) – Comparing Standard and Centered Results

Term	VIF	Term – Centered	VIF – Centered	Df

spi_comp	63.775125	cen_spi_comp	2.890851	1
spi_comp^2	68.752326	cen_spi_comp^2	1.360307	1
di_score	41.153273	cen_di_score	2.681510	1
di_score^2	46.657004	cen_di_score^2	1.809707	1
log_gdppc	182.168658	cen_log_gdppc	2.661729	1
log_gdppc^2	194.469997	cen_log_gdppc^2	1.464490	1

Note. Recall a VIF score greater than 10 is severely correlated and is problematic for interpreting regression coefficients. *factor(year)* is excluded from this analysis, which eliminated the need for $GVIF^{1/(2*\mathrm{Df})}$ estimation for categorical variables.

To reiterate, all predictors are statistically significant; this validates Ramsey RESET test alternative hypothesis that non-linearity exists. Nevertheless, despite correcting for multicollinearity and applying terms that significantly capture the complexity of the data, further analysis necessitates variable reductions to circumvent the possible risk of overfitting.

Further validated by AIC/BIC results, the quadratic model has lower AIC and BIC values compared to the linear model, indicating a better fit. The AIC for the linear baseline model is 7446.06, while the quadratic model's AIC is 7193.846. Similarly, the BIC for the linear model is 7671.324, and for the quadratic model, it is 7480.546. These scores steadily decrease after every additional quadratic term that enters the linear-baseline model.

Table X

Summary Results of Quadratic Fit Statistics: AIC/BIC Results, Adjusted R^2 & Net Change

Model	Quadratic Terms	AIC, BIC	Adj. R^2	Absolute Increase	Marginal Increase
Linear-Baseline	–	7446.06, 7671.324	0.7757	–	–
1 Quadratic	log_gdppc^2	7266.005, 7511.748	0.8072	+0.0315	+0.0315
2 Quadratic (A)	log_gdppc^2, spi_comp^2	7228.738, 7494.959	0.814	+0.0383	+0.0068
2 Quadratic (B)	log_gdppc^2, di_score^2	7211.045, 7477.266	0.817	+ 0.0413	+0.0098
3 Quadratic	log_gdppc^2, spi_comp^2, di_score^2	7193.846, 7480.546	0.8202	+0.0445	+0.0062

Note. Absolute Increase is calculated as the difference between each model's Adj. R^2 and the Linear-Baseline's Adj. R^2. *Marginal Increase* is the difference in Adj. R^2 compared to the model with one fewer quadratic term. 2 Quadratic (A) and (B)'s marginal increase is calculated based on the 1 Quadratic model's Adj. R^2.

Test 3: Moderators / Interactions

Test 2c assesses statistical significance of moderators, specifically: GNI Classifications (categorical), regime type (binary), regime type (continuous), EIU democracy score (continuous), regime change (democratization episode) and regime change (autocratization episode). ~~The results serve to validate ___'s examination, where significant differences in statistical capacity and SDG relationships were observed depending on GNI classifications.~~

Furthermore, the strength of interaction terms also informs the specific stratification strategy, that is, how this analysis groups countries for subgroup analysis – Test 3. Test 2c deals with the following sub-hypotheses:

- ***H0**:* Null – No interactions found
- ***H1d**:* The impact of SPI on SDG performance varies depending on GNI Classification [two-tailed]
- ***H1e**:* The impact of SPI on SDG performance varies depending on regime type: observed as a binary, categorical and continuous measure
- ***H1f**:* The impact of SPI on SDG performance varies depending on whether a country is experiences episodes of regime change: autocratization and democratization

$$SDG\ Performance = a + \beta_1 SPI + \beta_2 Z_{Predictor} + \beta_3 (SPI \times Z_{Moderator}) + \beta_n Controls + \epsilon$$

$$Main\ Effect\text{: } X(Z_{Moderator}),\ \text{when } Z = 0$$

With integrating binary and/or categorical moderators, β_2 (the coefficient for the moderator variable) represents the average difference of a given category from the baseline (reference category); the slope of $Z_{Predictor}$ is the effect, net of the baseline effect, of a given moderator (Z) on the response variable ($SDG\ Performance_{Overall}$).

Test 3 Results

Table X

Pooled OLS Interaction Models

	Dependent Variable: SDG Performance (Composite)				
	Moderator Variables (Z):				
	GNI Class	Regime-Type (Binary)	Regime-Type (Categorical)	DI Score	Regime Change Episode
	(1)	(2)	(3)	(4)	(5)
cen_spi_comp	0.176^{**}	0.260^{***}	0.205^{***}	0.306^{***}	0.293^{***}
	(0.079)	(0.046)	(0.077)	(0.032)	(0.038)
factor(income_level_recoded)1	4.081^{**}				
	(1.822)				
factor(income_level_recoded)2	6.757^{***}				
	(2.622)				
factor(income_level_recoded)3	5.656^{*}				

	(3.077)				
factor(regime_type_binary)1		0.845			
		(0.929)			
factor(regime_type_categ)1			1.411		
			(1.483)		
factor(regime_type_categ)2			1.943		
			(1.686)		
factor(regime_type_categ)3			3.035		
			(2.020)		
cen_di_score	0.072	0.172	-0.005	0.334	
	(0.223)	(0.307)	(0.327)	(0.249)	
aut_ep1					-0.910
					(0.657)
dem_ep1					0.616
					(0.730)
cen_log_gdppc	1.935**	3.230***	3.122***	2.976***	3.448***
	(0.774)	(0.404)	(0.399)	(0.386)	(0.390)
I(cen_log_gdppc2)	-0.690**	-0.843***	-0.995***	-1.026***	-0.778***
	(0.268)	(0.191)	(0.218)	(0.211)	(0.189)
cen_spi_comp * factor(income_level_recoded)1	0.188** (0.090)				
cen_spi_comp * factor(income_level_recoded)2	-0.013 (0.082)				
cen_spi_comp * factor(income_level_recoded)3	0.291*** (0.084)				
cen_spi_comp * factor(regime_type_binary)1		0.056 (0.050)			

cen_spi_comp * factor(regime_type_categ)1			0.080 (0.079)		
cen_spi_comp * factor(regime_type_categ)2			0.069 (0.081)		
cen_spi_comp * factor(regime_type_categ)3			0.159^{*} (0.087)		
cen_spi_comp * cen_di_score				0.030^{**} (0.012)	
cen_spi_comp * aut_ep1					0.040 (0.041)
cen_spi_comp * dem_ep1					0.035 (0.052)
Intercept	64.007^{***} (2.489)	69.072^{***} (0.777)	67.878^{***} (1.537)	69.473^{***} (0.577)	69.989^{***} (0.585)
Observations	1,236	1,236	1,236	1,236	1,261
R^2	0.842	0.811	0.816	0.816	0.813
Adjusted R^2	0.840	0.809	0.813	0.815	0.811

Note: All continuous variables are means-centered for multi-collinearity; all models apply robust standard errors for heteroskedasticity $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Interaction 1

HA: The impact of SPI on SDG performance varies depending on GNI Classification (income_level_recoded: 1 = Low; 2 = Lower-Middle; 3 = Upper-Middle; 4 = High)

The results indicate a partial moderation effect of Z (*income_level_recoded*) on the relationship between SPI score (*cen_spi_comp*) and SDG performance (*sdg_overall*). Specifically, interaction between SPI and SDG performance is significant for certain GNI country classifications – lower-middle-income and high-income – whereas countries grouped within upper-middle income do not show a significant moderating effect.

For countries grouped as *low-income* (reference group), a one-unit increase in the SPI score is associated with a 0.176-point increase (main effect) in the SDG overall score ($p < 0.05$*), holding all other variables constant. Provided that *low-income* is the reference category ($Z_{Level} = 0$), a one-unit increase in the SPI score is then associated with an additional 0.1883 (interaction effect) increase in SDG performance for lower-middle income countries from the initial main effect of 0.1757. This difference is statistically significant from ($p < 0.05$*), and the impact of SPI amounts to a total effect of 0.3639; if 0.1757 (main effect) + 0.1883 (interaction effect) = 0.3639.

Similarly, based on the initial main effect of 0.1757, a one-unit increase in the SPI score is associated with an additional 0.2914 (interaction effect) increase in SDG performance for high-income countries. This difference is highly statistically significant ($p < 0.001$***), and the impact of SPI amounts to a total effect of 0.4670; if 0.17565 (main effect) + 0.29137 (interaction effect) = 0.4670.

The only GNI country category with sufficient evidence against the alternative hypothesis, specifically that *an interaction exists and moderates the relationship between the main X (SPI) and Y (SDG Performance)*, is the Upper-Middle Income group classification. [small synthesis/analysis of results]

Interaction 2

HA: The impact of SPI on SDG performance varies depending on Regime Type (Binary: 0 = Autocracy; 1 = Democracy)

Provided that *autocratic regime* ($Z_{level} = 0$) is the reference category, a one-unit increase in the SPI score is then associated with an additional 0.0561 (interaction effect) increase in SDG performance for democratic regimes from the initial main effect of 0.2601. Although the main effect from the reference category, when $Z_{level} = 0$, is statistically significant, the net-difference is not statistically significant ($p > 0.05$). This suggests that being a democracy does not impact the relationship between SPI and SDG performance. In fact, the main effect being significant further suggests that the effect of SPI on SDG performance would happen regardless of the particular regime type.

Interaction 3

HA: The impact of SPI on SDG performance varies depending on Regime Type (Categorical: 0 = Closed Autocracy; 1 = Electoral Autocracy; 2 = Electoral Democracy; 3 = Full Democracy)

Like the binary measure (*regime_type_binary*), a more disaggregated grouping of regime-types (*regime_type_categ*) does not yield significant differences in the magnitude of impact that SPI has on SDG performance by subgroup. Results indicate that SPI's effect on SDG performance would occur regardless of regime-type. However, if there is a marginal readjustment of the alpha-confidence level ($a = 0.10$), there *would* be evidence to suggest that countries specifically classified as 'full democracies', amplify SPI's effect on SDG performance. Even so, further examination by applying alternative panel techniques and diagnostics checks could provide greater assessment of this particular interaction between described regime-type on the SPI to SDG relationship.

Interaction 4

HA: The impact of SPI on SDG performance varies depending on Democracy Score (cen_di_score: 0-1, continuous)

Coefficient(X): The effect of X on Y, when Z = 0 (centered to the *mean value*).
The positive and significant interaction term (*Z*: 0.0299, $p = 0.0138$) indicates that the effect of SPI on SDG performance increases in magnitude by about 0.03 points for every additional unit

increase in DI levels. This suggests a synergistic relationship, where countries with higher levels of democracy (proxied as DI score), are more likely to experience greater improvements in sustainable development directly from SPI compared to less-democratic countries.

Interaction 5

The results indicate that SPI affects SDG progress regardless of whether countries are experiencing episodes of autocratization (*aut_ep*) or democratization (*dem_ep*). Neither binary moderator of regime change, whether in a positive ($Z_{Democratizing}$) or negative ($Z_{Backsliding}$) directional status, interact with the relationship between statistical performance and sustainable development.

Although the results show no indication, it is important to emphasize that this is a pooled OLS regression model that sets all observations on equal footing despite our data's panel structure. With total observations constituted by over 150 countries over several years, this model violates the assumption of independence. Additionally, literature (Ştefanachi et.al, 20XX) and our previous diagnostics of non-linearity (see Test 2a and 2b results) suggest significant non-linearity in all essential predictors of SDG performance (i.e., DI, SPI and Log($GDP_{Per\text{-}Capita}$)).

COMPONENT 2: PANEL MEDIATION ANALYSIS

Panel Models – Motivation:

It would be impossible to control for all unobserved factors that could influence SDG scores and statistical capacity. Nevertheless, a key advantage of applying panel techniques is the to control for *time-invariant factors*, which include immutable country characteristics like geography and regional-cultures, which do not change over time (at least within the 2017-2023 time period). Component 2 employs first difference and fixed effects methods to secure a more robust analysis of our key variables: democratic governance, statistical capacity, and sustainable development. Additionally, said models provide a solid means to capture and analyze movement over time, and assess if/how variables shift mutually over time.

Fixed Effects – Motivation:

- A two-way fixed effects model eliminates

First Difference – Motivation:

Model Requisites

A fixed effects model requires at least two years of estimates on the same variable. At the same time, a first difference model requires consecutive country-year observations to calculate immediate changes. The range of years was strategically selected for the years 2016 to 2023 to circumvent missing data in years prior to the development of the Statistical Performance Indicators (SPI). This also allowed for a more balanced panel model, which contains all values of all variables in all countries across all specified years. This time-period conveniently reduced the need for imputation, which could underestimate statistical variance and statistical power.

Anticipating Reverse Causality

- Asking the question "how does one variable x affect variable y", we already assume that an effect between one variable and another indeed exists. We answered this question in section ___ of the paper (see Section X).
- However, in this statement, we are also making an assumption about directionality, specifically: X → Y.
- In order to cover our bases, it is important to consider the counter argument: could it have been that variable y affected variable x: X ← Y?

Technique:

- Panel Granger Tests
- **https://journals.sagepub.com/doi/pdf/10.1177/0049124119882473 (AMAZING THING TO CITE FOR METHODOLOGY)**

Results

I devised a series of models, each testing a unique variation and arrangement of predictor(s), including alternative linear/non-linear functional forms, in order to assess the most optimal model specification. Several statistical checks were adopted from Component 1 to ensure robust estimates, including:

COMPONENT 3: TREND ANALYSIS & REGRESSING REGIMES – CASE STUDIES [SAVE FOR LATER TIME]

COMPONENT 3: RESULTS & ANALYSIS [SAVE FOR LATER TIME]

DISCUSSION: LIMITATIONS & OPPORTUNITIES

Key limitations:

- COVID 19
- There are certainly other measures of statistical capacity worth considering such as Open Data Inventory (ODIN), and Information Capacity measures, Open Data Barometer (ODB), etc. (see appendix X).
- There are other measures of democracy/authoritarianism: The EIU measure serves as a starting point to potential future research involving other measures of democracy not limited to, Freedom House, Polity5, Varieties of Democracy, etc. (see appendix X)
- Interlinkage of the UN SDGs; the goals are designed to support each other. According to _____, … they're connected. Is it likely that an affect on one of the 17 SDGs is also attributed to an effect on the other 17 SDGs.
- Opportunity: Network visualization combined with panel lag models such as fixed effects can help understand how the 17 goals interact with each other and how an effect on one consequently affects another goal.
- Did not standardize OLS dependent or independent variables in analysis of statistical capacity measures. It was not necessary given that all indices were standardized to the standard normal.
- I left out important variables given their high correlation with other independent variables. For instance, countries' total population might have an impact on countries ability to capture accurate estimates on their population. Emerging economies with higher populations may not have the infrastructure to disseminate national census surveys as effectively as others with lower populations. It could be that countries with higher populations face greater budgetary constraints on policy research and/or statistical agencies, having to allocate a greater expenditure to accommodate a high population. That said, total population is highly correlated with GDP Per Capita. For interpretability, this study removes population and sustains GDP Per Capita, which captures aspects of the population and financial wellbeing.
- A next step for future analysis would, thus, integrate statistical techniques such as Principal Component Analysis (PCA) and/or other aggregation techniques to circumvent multicollinearity of predictor variables within the same model.

~~FOR ME: What Do I Mean by Statistical Capacity~~

- ~~I guess I am referring to the openness of data, the accessibility of data, whether data is taken down from official websites (during transition), whether we see less funding into research and development/statistics, whether the government is manipulating data to present what they want us to see/know (censorship), underreporting of certain sectors/issues, how the truth is being bended due to data malpractices…~~
- ~~The impact that all of this has on policy-making (across all policy sectors represented by the UN SDGs)~~

CONCLUSION

- …

GLOSSARY OF TERMS AND ACRONYMS

Statistical Capacity

- *Statistical Capacity:* "the ability of a country's national statistical system, its organizations and individuals to collect, produce, analyze and disseminate high-quality and reliable statistics and data to meet users' needs" (PARIS 21, 2018a, p. 9).
- Statistical Capacity Development: "The process by which individuals, organizations, and societies obtain, strengthen, and maintain capabilities to collect, produce, analyze, and disseminate high-quality and reliable statistics and data to meet users' needs to set and achieve their own development objectives. Thus, a nation's statistical capacity is directly linked to the ability to set and achieve development objectives that are country- or locally-owned." (M. Tichenor, 2022, p. 546)

UN Sustainable Development Goals (UN SDGs)

- UN SDGs: designed to serve as a universal and flexible policy blueprint with the sole purpose of promoting people, prosperity, planet, peace, and partnership – the '5 Ps'. The UN SDGs comprise of 17 global goals, 169 targets, and 244 indicators (and many more under review) created to facilitate progress towards achieving the 2030 Agenda.
- *The 2030 Agenda:*

Regime Type & Regime Transition

- Regime change: This study adopts Cambridge Dictionary's definition, specifically "a complete change of government, especially one brought about by force" (Dictionary).
- Regime Type:
- Democracy: "Extent to which citizens can choose their political leaders in free and fair elections, enjoy civil liberties, prefer democracy over other political systems, can and do participate in politics, and have a functioning government that acts on their behalf." – Economist Intelligence Unit.
- Autocracy:

Democratization & Autocratization

- Autocratization: Varieties of Democracies accepts the definition of autocratization, commonly known as 'democratic backsliding', as "any movement towards autocracy which starts within democracies or autocracies" (cf. Maerz et al., 2023; Lührmann and Lindberg, 2019). Democratic backsliding is thereby assumed synonymous with the term 'autocratization' and will used interchangeably over the course of this study.
- Democratization: This study accepts Wilson et. al conceptualization of democratization described as "any movement towards democracy which starts in autocracies or democracies" (2022). This definition is similarly accepted by Varieties of Democracies.
- Democratic backsliding is assumed synonymous with the terms 'autocratization' and 'democratic regression.'

Research Terminology & Abbreviations

Statistical Performance Indicators (SPI):

- Statistical Performance Index: composite score of all 5 SPI pillars
- For research purposes, SPI refers to the overall score (Statistical Performance Index)
- Statistical Performance Index = SPI = *spi_comp*

Statistical Capacity Indicators (SCI):

- Statistical Capacity Index = composite score of SCI disaggregated dimensions
- For research purposes, SCI refers to the overall score – Statistical Capacity Index
- Statistical Capacity Index = SCI = *sci_overall*

Other Abbreviations

- EIU Democracy Index = DI = *di_score*
- Log of GDP per-capita = *log_gdppc*

Acknowledging views/perspectives of labeling countries:

If You Shouldn't Call It The Third World, What Should You Call It? (article link)

- Claims: "Developing Countries" is offensive and projects westernized conceptions about global order and ideal human
- Theres other ways to call 'developing' countries:
    - Global South
    - low income, lower middle income, middle income and high income (WHO)
    - Majority World

APPENDIX A:

UN SDG Indicators & SDG Index – Notes

Sustainable Development Goals (SDGs):
For clarification, all the 17 goals are measurable benchmarks constituted by over 250 indicators (many of which are still under review). The SDG Index comprises of 98 global indicators that contribute to an overall composite score, which is the main response variable of this study (Appendix B).

Link: https://dashboards.sdgindex.org/explorer?metric=overall

SDG 17, Targets 18 & 19
Statistical capacity helps constitute SDG 17 – "Strengthen the means of implementation and revitalize the global partnership for sustainable development."

Target 17.18: By 2020, enhance capacity-building support to developing countries, including for *least developed countries and small island developing States, to increase significantly the availability of high-quality, timely and reliable data disaggregated by income, gender, age, race, ethnicity, migratory status, disability, geographic location and other characteristics relevant in national contexts*

- **Indicator 17.18.1: Statistical capacity indicators**
- Indicator 17.18.2: Number of countries that have national statistical legislation that complies with the Fundamental Principles of Official Statistics
- Indicator 17.18.3: Number of countries with a national statistical plan that is fully funded and under implementation, by source of funding

Target 17.19: By 2030, build on existing initiatives to develop measurements of progress on sustainable development that complement gross domestic product, and support statistical capacity-building in developing countries

- Indicator 17.19.1: Dollar value of all resources made available to strengthen **statistical capacity** in developing countries
- Indicator 17.19.2: Proportion of countries that (a) have conducted at least one population and housing census in the last 10 years; and (b) have achieved 100 percent birth registration and 80 percent death registration

Figure A

The United Nations Sustainable Development Goals (SDGs)

Source: *United Nations Department of Global Communication, 2019*

APPENDIX B:

Additional Notes on Variables Used & Methodologies

B.1) SDG Index, Sustainable Development Report (2024) – Methodology

Link to Methodology (online): https://dashboards.sdgindex.org/chapters/methodology

- Only for countries with 80% of data
- Many data points have not been able to be updated since Feb 2022, current 2024 estimates may not take into account large shocks like the war in Ukraine
- The 2024 SDG Index covers 167 countries
- Includes 98 global indicators (the best we have so far).

Applicable only to countries with 80% of data availability. Many data points have not been updated since February 2022, and current 2024 estimates may not reflect significant events such as the war in Ukraine. The 2024 SDG Index includes 167 countries and features 98 global indicators (the best they have so far).

[Maybe include in abstract with link to webpage] Their 5 criteria in selecting indicators:

- 1. Their global relevance and applicability to a broad range of country settings.
- 2. Statistical adequacy: The indicators represent valid and reliable measures.
- 3. Timeliness: The indicators are current and published on a timely schedule.
- 4. Coverage: Data is available for at least 80 percent of UN member states with a population > 1 million.1 [WAIT, so this means that we can't predict decreases in SDG estimates from decrease in SPI, specifically in countries with missing data?]
- 5. Distance to targets must be measurable (optimal performance can be defined).

Where they get indicator data:

- They use official SDG indicators by the UN Statistical Commission, where possible. If there are gaps in the data, they collect from official and unofficial providers.
- ⅔ of data: “Most of the data (around two-thirds) come from international organizations (World Bank, OECD, WHO, FAO, ILO, UNICEF, other) which have extensive and rigorous data validation processes.”
- ⅓ of data: “Other data sources (around one-third) come from less traditional statistics including household surveys (Gallup World Poll), civil society organizations and networks (Oxfam, the Tax Justice Network, the World Justice Project, Reporters Without Borders), peer-reviewed journals (e.g., to track international spillovers), and geographic information systems (GIS)”
- Refer to Table A.5 in the Sustainable Development Report: Methodology for an extensive list of indicator sources.

Limitations

- “Due to changes in the indicators and refinements in the methodology, SDG Index rankings and scores from one edition cannot be compared with the results from previous editions.”

    - That said, SDG Index estimates are calculated retroactively using 2024 measurements and specifications.
- They "do not incorporate estimates received directly from national statistical offices. Data providers may adjust national data to ensure international comparability"
    - This is good because it reduces biases from self-reporting
- Long validation process from international organizations (where data for SDG Index is traced) may cause delays in publishing, so SDG estimates might differ from more recent data from countries

Missing data and imputation
- They don't impute for missing data except for very few and exceptional circumstances. Please refer to their codebook online to see the imputed data.

Method for constructing SDG Index
This specific link to section:
https://dashboards.sdgindex.org/chapters/methodology#c-method-for-constructing-the-sdg-index-and-dashboards

Three steps:
- 1. Establish performance threshold (to know what is good and bad performance- upper and lower bounds)
- 2. Normalize all indicators to ensure comparability
- 3. Aggregate all indicators that are within and across the 17 SDGs

See all three steps in section C. of their methodology.

Normalization
- Representative quote: "all rescaled variables were expressed as ascending variables (i.e., higher values denoted better performance). In this way, the rescaled data became easy to interpret and compare across all indicators: a country that scores 50 on a variable is halfway towards achieving the optimum value, whereas a country with a score of 75 has covered three quarters of the distance from worst to best."

Weighing and aggregation:
- No clear consensus by experts
- Representative quote: "As a normative assumption, we therefore opted for fixed, equal weight to every SDG to reflect policymakers' commitment **to treat all SDGs equally and as an integrated and indivisible set of goals.** This implies that to improve their SDG Index score countries need to place attention on all goals with a particular focus on goals where they are furthest from achieving the SDGs and where incremental progress might therefore be expected to be fastest."
- https://dashboards.sdgindex.org/chapters/methodology#table-a-4-spillover-indicators-and-categories

B.2) Statistical Performance Indicators (SPI)

SPI Pillars (5)

Utilized as both an effect and a mediator variable in this study, the Statistical Performance Index (SPI) is characterized as a measure of the "performance of national statistical systems," and comprises of the following five pillars:

"

I. Data Use: Statistics have value only if they are used. So the first pillar is data use. A successful statistical system produces data that are used widely and frequently.

II. Data Services: A range of services connects data users to producers and facilitate dialogues between them, thus building trust and a sense of value.

III. Data Products: The dialogues between users and producers drive the design and range of statistical products and their accuracy, timeliness, frequency, comparability, and levels of disaggregation. The products signal whether countries are able to produce indicators related to the 17 Sustainable Development Goals.

IV. Data Sources: To create useful products, the statistical system needs to draw on sources inside and outside the government. Data collection thus goes beyond the typical censuses and surveys to include administrative and geospatial data as well as data generated by private firms and citizens.

V. Data Infrastructure: A mature statistical system has well-developed hard infrastructure (legislation, governance, standards) and soft infrastructure (skills, partnerships) as well as the financial resources to deliver useful—and widely used—data products and services.

"

Table B

Breakdown of SPI Dimensions by Measurability

SPI Pillar	Developed Methodology	Lack of Methodology
Data Use (1)	1.1 International Bodies	1.2 Legislature 1.3 Executive 1.4 Civil Society 1.5 Academia
Data Services (2)	2.1 Quality of Data Releases 2.2 Richness & Openness of Online Access 2.3 Availability & Use of Data Services	2.4 Effectiveness of Advisory & Analytical Services Related to Statistics
Data Products (3)	3.1 Social (SDG 1-6) 3.2 Economic (SDG 7-12) 3.3 Environmental (SDG 13-15) 3.4 Institution (SDG 16-17)	-
Data Sources (4)	4.1 Statistical Office (Censuses & Surveys)	4.4 Private Sector Data/Citizen Generated Data

	4.2 Administrative Data 4.3 Geospatial Data	
Data Infrastructure (5)	5.1 Legislation & Governance 5.2 Standards & Methods 5.3 Finance (Domestically & From Donors)	5.4 Skills 5.5 Partnership
Frequency of all SPI dimensions	14 / 22 dimensions, 51 measurable indicators	8 / 22 dimensions

Note. [a.] The Developed Methodology columns lists dimensions for which maintain the necessary data sources and methodological rigor for measurement. [b.] The Lack of Methodology column tracks dimensions that lack said components. [c.] Numbers are assigned purely for organizational purposes and do not follow the labeling scheme of the SPI-dimention framework.

Builders of the SPI (2021) note that eight of the 22 dimensions listed do not have an indicator with a developed methodology or have incomplete data collection (p.11). Available dimensions (14) make up a grand total of 51 measurable indicators that collectively contribute to an overall SPI score (Dang et. al, 2021). Table B provides a breakdown of all dimensions

~~B.3) Statistical Capacity Indicators (SCI)~~

~~"The SPI framework covers several of the same attributes as the SCI, such as statistical methodology, data, and periodicity, but expands into new areas as well" (Serajuddin, p. 6)~~

B.4) Democracy Index, Economist Intelligence Unit (EIU DI)

…

B.5) Alternative Measures [probably not necessary]

Statistical / Reporting / Data Capacity Measures:

- Open Data Inventory (ODIN), Open Data Watch: https://odin.opendatawatch.com/
- Open Data Barometer (ODB): https://opendatabarometer.org/?_year=2017&indicator=ODB
- Information Capacity Index: https://datafinder.qog.gu.se/dataset/icd#:~:text=Description:,polities%20from%201750%20to%202015.

Democracy Measures:

- Polity5, Polity Project:
- Freedom House:
- Varieties of Democracy: _____

Controls

- State capacity
- R&D spending
- state_fragility (state fragility index)

SDG Development Measure:

- Global State of Democracy (GSoD) Indices. The Sustainable Development Goals and the GSoD Indices, Revised Edition (2023) (Article Link)
    - "The GSoD Indices measure democratic performance in 173 countries and provide data to track progress on 7 SDGs (1, 2, 3, 4, 5, 10 and 16)."

APPENDIX C:

Codebook of All Variables

https://docs.google.com/spreadsheets/d/16Mt96CjJSWWkS9rS8tqkUh1G9b8qZDdbJTywtuB89tA/edit?gid=7874732#gid=7874732

Statistical Performance Indicators

- SPI Variables: Statistical Performance Indicators: Series Code [Categorical] = Overall Score (IQ.SPI.OVRL), Data Use (IQ.SPI.PIL1), Data Services (IQ.SPI.PIL2), Data Products (IQ.SPI.PIL3), Data Sources (IQ.SPI.PIL4), and Data Infrastructure (IQ.SPI.PIL5).
- SCI Variables: Statistical Capacity score Indicator Code [Categorical] = Overall average (IQ.SCI.OVRL).
- Merge Variable: Country Code (country_code); Series Code-Indicator code; Year [2000-2023].

Varieties of Democracy – FULL & ERT DATA

- The (reg_change) variable is an ordinal categorical variable [-1, 0, 1] that denotes the direction to which regimes have transitioned (e.g., toward democracy or toward autocracy).
- ERT_V-Dem Variables: Regimes of the World (v2x_regime) [categorical: 0-3]; Regime transition (reg_trans) [categorical: -1,0,1]; Regime type (reg_type) [binary: 0,1]; etc…
- Full_V-Dem Variables: Electoral democracy index (v2x_polyarchy); Liberal democracy index (v2x_libdem); Participatory democracy index (v2x_partipdem); Deliberative democracy index (v2x_delibdem); Egalitarian democracy index (v2x_egaldem)
- Merger Variables: Country ID (country_id); Year (year) [integer: 2000-2023]; Country Code (country_code)

APPENDIX D:

Descriptive Statistics

Figure D1

Descriptive Statistics of Statistical Capacity Measures

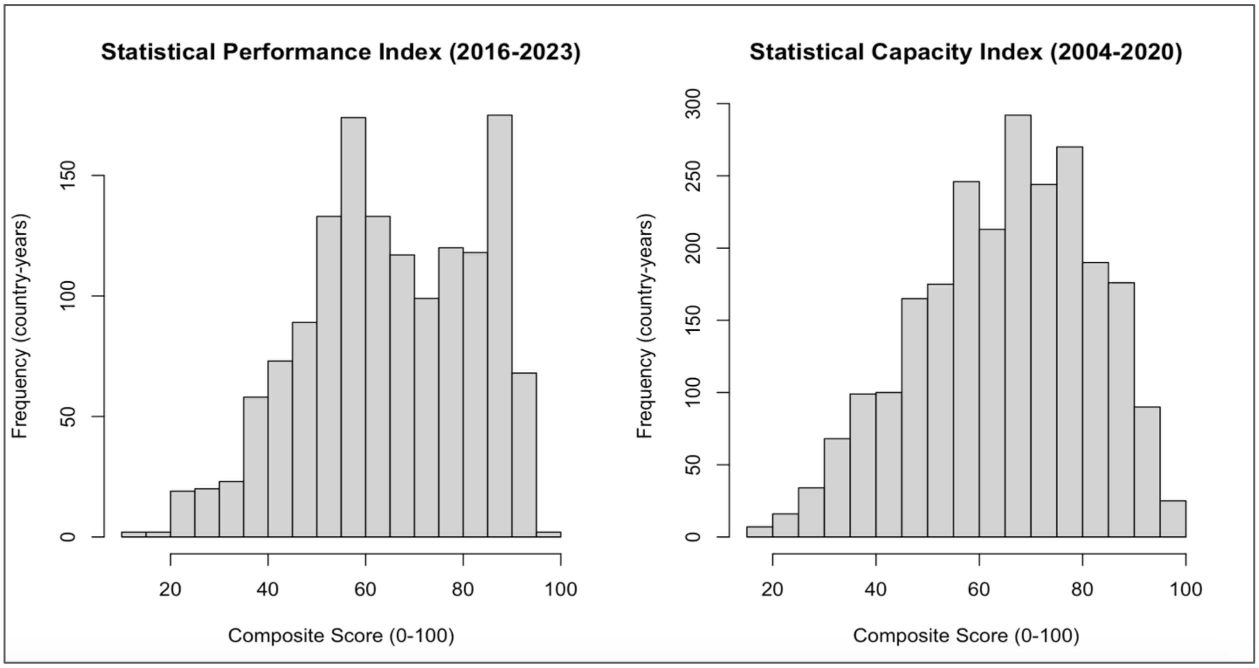
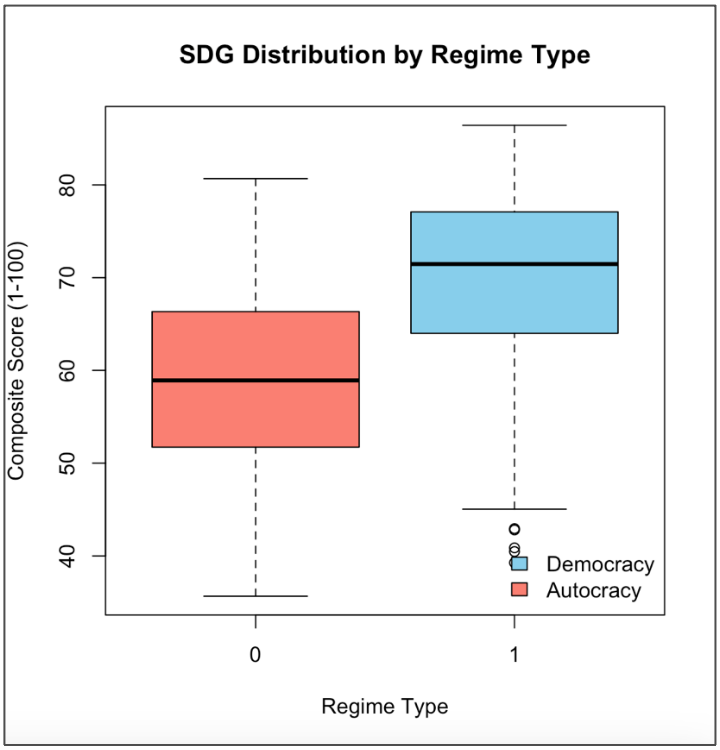


Figure D2

Box plots Comparing Autocracies and Democracies by SDG Scores



The composite score of UN SDG progress, an aggregation of performance scores 17 global objectives, appears associated with regime type. Specifically, there have been more instance of democracies exhibiting high SDG progress compared to autocracies.

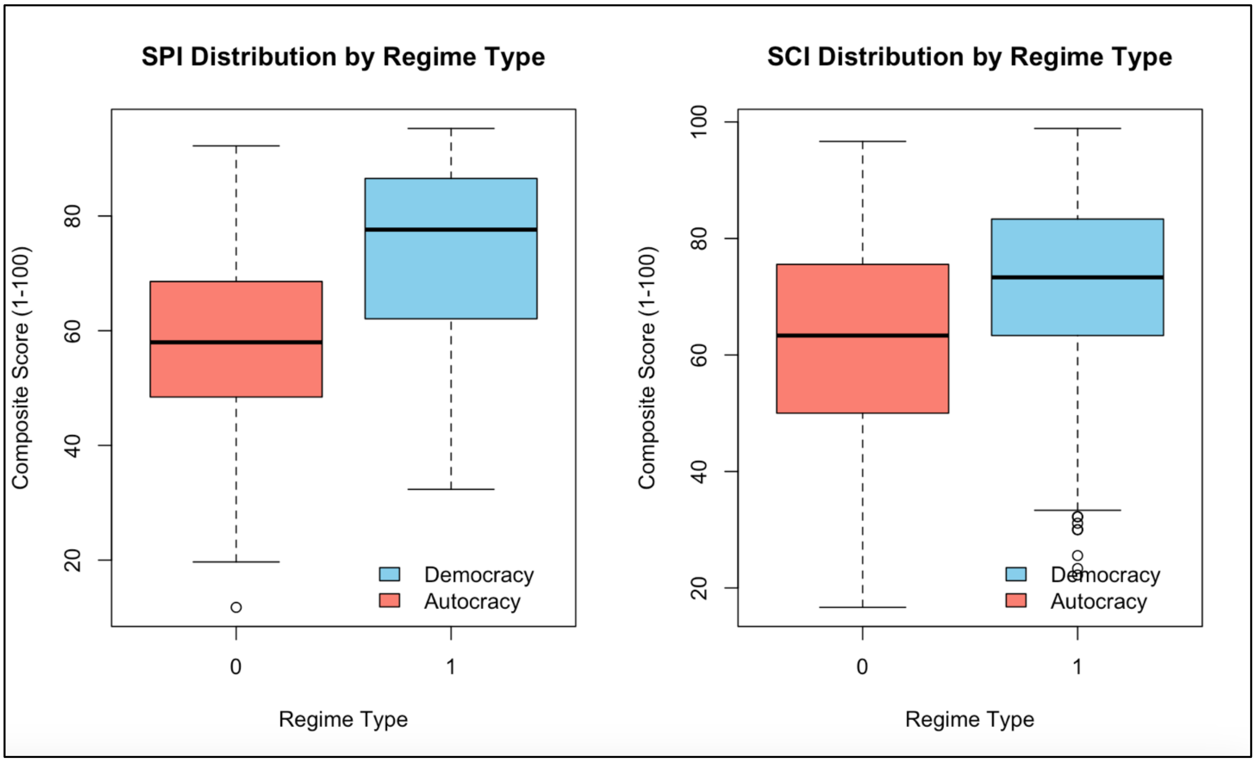
Table D1

Distribution of SDG Composite Scores by Regime Type (counts)

regime_type_2	Q1 (Lowest)	Q2	Q3	Q4 (Highest)	NA
0	639	481	357	72	265
1	174	314	468	768	193
NA	27	45	15	0	801

Figure D3

Distribution of Statistical Capacity Measures by Regime Type



- Regardless of which measure, we see that statistical capacity has a stronger association with democracies than autocracies.

Table D2

Distribution of countries' SPI Composite Scores by Regime Type (counts)

regime_type_2	Q1 (Lowest)	Q2	Q3	Q4 (Highest)	NA
0	212	159	80	102	1261
1	108	109	76	231	1393
NA	57	24	1	0	806

Table D3

Distribution of SCI Composite Scores by Regime Type (counts)

regime_type_2	Q1 (Lowest)	Q2	Q3	Q4 (Highest)	NA
0	335	227	217	278	757
1	118	124	199	323	1153
NA	120	62	15	5	686

- Frequency distribution of countries by regime-type classifications [histogram]
- Number of countries that experienced regime change since 2000
    - Number experiencing democratization since 2000

- Number experiencing autocratization since 2000

APPENDIX E

Preliminary Analysis – Statistical Capacity and SDG Performance Relationship

A preliminary assessment of the basic factors that establish a foundation for further study of statistical capacity's impact on SDG performance is provided here. Component 1 of this study necessitates initial assumptions about the relationship between the main dependent variable (i.e. SDG performance) and select statistical capacity variables (i.e., Statistical Performance Index (SPI) and Statistical Capacity Index (SCI)). Additional calculations are done to assess the relationship between democracy/regime-type – via EIU's Democracy Index (DI) – and two dependent variables statistical capacity (1) and SDG performance (2), which is investigated at length in component 2 of this study. This preliminary investigation, supported alongside existing literature, applies the following statistical methods for analysis: correlation coefficients, R^2 estimates, and naïve ordinary least squared models. For reference, this analysis is structured

H0: No relationship
H1: Statistical capacity impacts SDG performance [two-tailed]
H2: Statistical capacity strongly and positively impacts SDG performance [one-tailed]

Table E1

Correlation Coefficients & R2 Estimates

Relationship	Correlation Coefficient	R^2	Observations (country-years)
SDG ~ SPI	0. 784880	0. 616037	1300
SDG ~ SCI	0. 646503	0. 417966	2055
SDG ~ DI	0. 672644	0. 452450	2544
SPI ~ DI	0.676171	0.4572067	1247
SCI ~ DI	0.477767	0.2282613	1521

Note. SDG = SDG composite score (1-100); SPI = SPI composite score (1-100); DI = DI composite score (1-100).

The results demonstrates that SDG performance (SDG) composite scores are most strongly associated with the Statistical Performance Index (SPI), which shows a high correlation coefficient (0.78) and explains about 62% of the variance in SDG scores. Both the Statistical Capacity Index (SCI) and the Democracy Index (DI) also exhibit moderate positive correlations with SDG scores (0.65 and 0.67, respectively), accounting for 42% and 45% of the variance. This could indicate that countries with higher statistical performance and stronger democratic institutions tend to achieve better SDG outcomes, with SPI emerging as the most influential predictor among the indices examined.

Both SPI and SCI are also positively correlated with the Democracy Index, with SPI (0.68) showing a stronger association than SCI (0.48). These findings highlight the interconnectedness of statistical capacity and governance quality, suggesting that improvements in national statistical systems and democratic governance are mutually reinforcing factors in advancing sustainable development. Still, a causal relationship undoubtably warrants further examination, much of which is explored in this paper.

The correlation analysis provides a useful overview of the relationships between these indices, but further analysis is needed to understand the causal mechanisms and the impact of these indices on SDG outcomes.

Naïve OLS Models: Statistical Capacity Measures

I. *ols_spi_naive*: $SDG\_Performance_{Overall} = \beta_0 + \beta_1 SPI_{Overall} + \epsilon$
II. *ols_sci_naive*: $SDG\_Performance_{Overall} = \beta_0 + \beta_1 SCI_{Overall} + \epsilon$
III. *ols_multiple_naive*: $SDG\_Performance_{Overall} = \beta_0 + \beta_1 SPI_{Overall} + \beta_2 SCI_{Overall} + \epsilon$

Table E2

Naïve OLS Models: SPI & SCI x SDG

	Dependent Variable: SDG Composite Score		
	SPI Only (1)	SCI Only (2)	SPI + SCI (3)
SPI Composite Score	0.4781^{***}		0.2878^{***}
	(0.0105)		(0.0337)
SCI Composite Score		0.3921^{***}	0.1531^{***}
		(0.0102)	(0.0323)
Intercept	34.9463^{***}	33.8819^{***}	35.8644^{***}
	(0.7206)	(0.7115)	(1.2774)
Observations	1,300	2,055	596
R^2	0.6160	0.4180	0.4651
Adjusted R^2	0.6157	0.4177	0.4633
Note:			$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Based on the data, the impact of SCI on SDG and SPI on SDG are statistically significant, in all Naïve Ordinary Least Squares models. In model I (*ols_spi_naive*), SPI appears to have a greater impact on SDGs (0.47806, p-value < 0.001) compared to that of SCI (0.39081, p-value < 0.001) in model II (*ols_sci_naive*). Similarly, in model III, with both predictors (*ols_multiple_naive*), SPI outperforms SCI in terms of their coefficient values (0.28779 > 0.15311). All predictors in across all three models are highly statistically significant (p-value < 0.001). To reiterate, these are preliminary models without controls other than that of statistical capacity found in model III. The corresponding estimates also don't account for multiple time periods of countries, or potential non-linearity. These concerns are addressed over the course of the study.

Naïve OLS Models: Impact of Democracy Levels on Dependent Variables

I. *ols_spi_di_naive:* $SPI_{Overall} = \beta_0 + \beta_1 DI_{Overall} + \epsilon$

II. *ols_sdg_di_naive:* $SDG\_Performance_{Overall} = \beta_0 + \beta_1 DI_{Overall} + \epsilon$

Table E3

Naive OLS Models: DI & SPI/SDG

	OLS Models:	
	SDG ~ DI	SPI ~ DI
	(1)	(2)
DI Composite Score	5.0122***	3.2230***
	(0.1548)	(0.0703)
Intercept	39.9418***	47.6892***
	(0.9185)	(0.4178)
Observations	1,247	2,544
R^2	0.4572	0.4524
Adjusted R^2	0.4568	0.4522

Note. SDG and SPI composite scores are the dependent variables for models SDG ~ DI (1) and SPI ~ DI (2), respectively. *p<0.1; **p<0.05; ***p<0.01

The naive OLS models indicate that the Democracy Index (DI) has a strong positive relationship with both the Statistical Performance Index (SPI) and the Sustainable Development Goals (SDG) composite scores. The coefficients suggest that a one-unit increase in DI is associated with a 5.0122 increase in SPI and a 3.22301 increase in SDG scores, both statistically significant at $p < 0.001$. This suggests that countries with higher democracy scores tend to have better statistical performance and SDG outcomes.

APPENDIX F

Evidence of Heteroskedasticity – Component 1 Models

The Breusch & Pagan's (1979) Breusch-Pagan Test for heteroskedasticity was utilized to examine whether residuals are constant across model observations, which signals unaccounted non-linear relationships; such would hint to non-linear macro factors such as GDP per-capita and omitted variables like population, which can cause misdiagnoses of model relationships ultimately biasing results. This is also important because Ordinary Least Squares (OLS) models assume constant error variance. In such a complex world of diverse cultural and ever-changing political structures across almost 200 countries, cross-national data, especially in development economics, are rarely ever constant or linear.

Table F

Breusch-Pagan Test Results

Model	Statistic	P-value	Df	Method
(1) ols_spi	206.5	7.05e-39	10	Studentized Breusch-Pagan
(2) ols_sci	86.6	4.22e-12	15	Studentized Breusch-Pagan
(3) ols_multiple	31.1	1.33e-4	8	Studentized Breusch-Pagan

Note. [blahblahblah something here]

The results in Table F indicate strong evidence of heteroskedasticity in all three models. The small p-values in all models indicates that the variance of residuals are not constant across observations in all three models. This reinforces the motivation behind applying robust standard errors, which have been integrated to all OLS models. Without Robust SEs, there is a risk of inflated t-statistics, leading to false significance and misinterpretation of results. Despite the improvement from 206.5 (*ols_spi*) and 86.6 (*ols_sci*) to 31.1 (both), there remains statically significant heteroskedasticity in the combined model. Both statistical capacity measures create a better-specified model (*ols_multiple*), though not enough to eliminate heteroskedasticity entirely. Ultimately, these results serve to support the inclusion robust standard errors into our models for better analysis.

APPENDIX G

Evidence of Functional Form Misspecification – Component 1 Models

Table G

RESET Test Results of Test 1 OLS Models

Model	Statistic	P-value
(1) ols_base	77.854647	1.946920e-82
(2) ols_sci	25.230029	1.192191e-28
(3) ols_multiple	8.873893	1.827852e-11

Note. ols_spi (model 1) from Test 1 is renamed to *ols_base* to indicate the most suitable baseline for continued analysis; Estimates are corrected for heteroskedasticity-robust clustered errors.

The Ramsey RESET test (Ramsey, 1969) for model misspecification highlights that all three models (*ols_base*, *ols_sci*, and *ols_multiple*) exhibit significant signs of misspecification. Specifically, the p-values for the RESET test are all less than 0.001, indicating strong evidence against the null hypothesis of no misspecification. This suggests that the models do not adequately capture the relationship between the predictors and the SDG composite scores due to the existence of non-linear or combined functional form factors. Such is not limited to omitted variables, non-linear or high order polynomial relationships, or other forms of model misspecification.

Given the high variability of statistical capacity measures and control variables like GDP per capita, misspecification is possibly the result of omitted interaction terms or heteroskedasticity (Appendix F).

References [CITATIONS TO BE INCLUDED]

Lührmann, A., & Lindberg, S. I. (2019). A third wave of autocratization is here: what is new about it? *Democratization*, *26*(7), 1095–1113.

Ramsey, J. B. (1969). Tests for specification errors in classical linear least-squares regression analysis. Journal of the Royal Statistical Society: Series B (Methodological), 31(2), 350–371.

Nelder, J. A. (1977). A Reformulation of Linear Models. *Royal Statistical Society. Journal. Series A: General*, *140*(1), 48–63. https://doi.org/10.2307/2344517

<u>EIU Data:</u>

"Data Page: Democracy index", part of the following publication: Bastian Herre, Lucas Rodés-Guirao, and Esteban Ortiz-Ospina (2013) - "Democracy". Data adapted from Economist Intelligence Unit, Gapminder. Retrieved from https://ourworldindata.org/grapher/democracy-index-eiu [online resource].

Economist Intelligence Unit. (2006-2024). Democracy index [dataset]. Our World in Data. Retrieved May 9, 2025, from https://ourworldindata.org/grapher/democracy-index-eiu.

Economist Intelligence Unit. (2021). Democracy Index 2021: The China challenge. Economist Intelligence Unit.

Economist Intelligence Unit. (2022). Democracy Index 2022: Frontline democracy and the battle for Ukraine. Economist Intelligence Unit.

Economist Intelligence Unit. (2023). Democracy Index 2023: Age of Conflict. Economist Intelligence Unit.

Economist Intelligence Unit. (2024). Democracy Index 2024: What's wrong with representative democracy? Economist Intelligence Unit.

Gapminder. Democracy Index v4 [original data]. Retrieved from https://ourworldindata.org/grapher/democracy-index-eiu.

<u>SDG Index Data</u>

Sachs, J.D., Lafortune, G., & Fuller, G. (2024). The SDGs and the UN Summit of the Future. Sustainable Development Report 2024. Paris: SDSN, Dublin: Dublin University Press. doi:10.25546/108572