



Composite versus Disaggregated Governance and Sustainable Development: A Cross-Country Panel Analysis of Multicollinearity and the Apparent Sign Reversal of Regulatory Quality (2016–2023)

Phan Nhan Trung*
Thu Dau Mot University, Viet Nam

Received: May 2, 2026

Revised : August 11, 2026

Accepted : September 14, 2026

Online : October 14, 2026

Abstract

Cross-country research on governance and sustainable development frequently draws on the Worldwide Governance Indicators (WGI) and increasingly seeks to identify the distinct contribution of regulatory quality. This study examines whether such disaggregation is statistically advisable using a country-year panel of 160 countries observed during 2016–2023 ($N = 1,161$). Two measurement strategies are compared on the same dataset: a composite governance index that averages four WGI dimensions (Strategy A) and a specification that enters the four dimensions individually (Strategy B). The four WGI dimensions are correlated between 0.71 and 0.94, and the variance inflation factors for Government Effectiveness, Rule of Law, and Regulatory Quality exceed conventional thresholds in Strategy B. Under Strategy A, the composite governance index is positively and significantly associated with Sustainable Development Goals (SDG) performance in pooled and random-effects specifications, consistent with prior literature. Under Strategy B, however, the pooled-OLS coefficient on Regulatory Quality reverses sign and is negatively associated with SDG performance, while Government Effectiveness retains a large positive coefficient. Quantile regressions indicate that the negative coefficient on Regulatory Quality is not robust across the SDG distribution and loses statistical significance at most quantiles, and the negative coefficients do not survive estimation with the Hausman-preferred fixed-effects model. The study interprets these patterns as a measurement artefact rather than as substantive evidence that regulatory quality reduces SDG performance, and recommends that cross-country analyses of governance and sustainability either rely on a composite governance index or formally diagnose multicollinearity before assigning substantive meaning to the coefficients of individual WGI dimensions. The contribution is offered as a methodological caution for associational, non-causal governance and sustainability measurement research, where dimension-specific claims about regulatory quality are increasingly common.

Keywords: *Regulatory Quality, Worldwide Governance Indicators, Multicollinearity, Measurement Risk, Sustainable Development Goals, Cross-Country Panel, Measurement Reliability*

INTRODUCTION

Cross-country research on governance and sustainable development has grown rapidly during the Sustainable Development Goals (SDG) era, with a particularly prominent role for the Worldwide Governance Indicators (WGI) developed by the World Bank (Kaufmann et al., 2011; World Bank, 2024). Among the WGI dimensions, regulatory quality has attracted growing attention because regulatory capacity intersects directly with risk management, compliance, and the implementation of sustainability policies (Glass & Newig, 2019; Biermann et al., 2022; Lafortune et al., 2024). A natural empirical step is therefore to disaggregate the WGI and report the individual contribution of regulatory quality to SDG performance, alongside that of voice and accountability, government effectiveness, and rule of law. The present study examines whether this disaggregation is statistically advisable.

The motivation for the study is methodological rather than theoretical. The four WGI dimensions used in most cross-country governance applications are perception-based composite



indices that draw on overlapping sources of expert assessment and survey data (Kaufmann et al., 2011). They are well-known to be highly intercorrelated. When several highly collinear regressors are entered jointly, ordinary least squares estimates remain unbiased in expectation, but individual coefficients become unstable, their standard errors inflate, and—particularly in the presence of suppressor effects—their signs can reverse relative to bivariate or composite-based estimates (Wooldridge, 2010; Baltagi, 2021). In applied governance research, such reversals are not always interpreted as a measurement artefact. They are sometimes treated as substantive findings and used to motivate claims that one WGI dimension is more or less important than another for sustainability outcomes. The risk of this kind of misinterpretation appears to have grown as governance and sustainability research increasingly aims at dimension-specific policy implications.

Against this background, the present study compares two measurement strategies on the same cross-country panel. Strategy A constructs a composite governance index as the row mean of the four WGI dimensions and enters this index as the main regressor. Strategy B enters the four dimensions separately. The empirical comparison is designed to make the consequences of measurement choice visible. Under Strategy A, the composite governance index displays the conventional positive association with SDG performance. Under Strategy B, regulatory quality reverses sign and becomes negatively associated with SDG performance, while government effectiveness retains a large positive coefficient. Variance inflation factors confirm that several WGI dimensions are above the conventional tolerance threshold under Strategy B. The negative coefficients on the individual dimensions are, however, unstable: they lose significance under quantile estimation and reverse under fixed effects, as documented in Section 4.

The argument advanced in this paper is that the negative coefficient on regulatory quality under Strategy B is most plausibly interpreted as a measurement artefact rather than as substantive evidence that better regulatory quality is associated with worse SDG performance. The interpretation is offered with appropriate caution: collinearity diagnostics are necessary but not sufficient to rule out genuine differences across WGI dimensions, and the present study does not claim to settle the underlying conceptual debate. Rather, it documents an empirical regularity that researchers and reviewers should consider when designing or evaluating governance–sustainability analyses, particularly those that emphasize dimension-specific findings about regulatory quality. Specifically, the study pursues three objectives: (i) to compare composite and disaggregated governance measures on an identical sample; (ii) to assess how multicollinearity affects the stability of the coefficients on individual WGI dimensions; and (iii) to evaluate the consequences for SDG-related panel inference. Earlier SDG-period assessments have consistently emphasized that progress is uneven and that governance conditions vary widely across regions (Sachs et al., 2022).

The contribution is positioned as a methodological note for the governance, risk, compliance, and sustainability literatures. From a risk management perspective, the analysis highlights a measurement risk embedded in the choice of how to operationalize WGI: an apparently neutral decision about whether to use a composite index or individual dimensions can shift the substantive conclusion of an empirical paper. From a compliance research perspective, the study cautions against treating coefficient estimates on individual WGI dimensions as direct measures of regulatory effectiveness, since they may reflect statistical redistribution rather than underlying institutional differences. The study therefore contributes a small but practical building block to the broader effort to align measurement practice with substantive interpretation in cross-country governance and sustainability research.

The remainder of the paper is organized as follows. Section 2 reviews the relevant literature on governance measurement and develops the hypotheses. Section 3 describes the data, variables, research framework, and econometric specification. Section 4 reports the empirical comparison of

Strategy A and Strategy B. Section 5 discusses the findings and their implications. Section 6 concludes, and Section 7 acknowledges limitations and identifies avenues for future research.

LITERATURE REVIEW

The WGI in cross-country governance research

The Worldwide Governance Indicators have become the most widely used cross-country measure of institutional quality. Each WGI dimension summarizes a set of underlying indicators drawn from expert assessments and survey instruments using an unobserved-components model (Kaufmann et al., 2011). The six dimensions—voice and accountability, political stability, government effectiveness, regulatory quality, rule of law, and control of corruption—are conceptually distinct but empirically share substantial common variation. The World Bank (2024) provides updated WGI data through recent years, supporting their continuing use in cross-country panel studies. Recent reviews note both the strengths of the WGI as a comparative measurement instrument and the limitations of perception-based composite indices for capturing on-the-ground implementation (Glass & Newig, 2019; Biermann et al., 2022).

Empirical applications use the WGI in two main ways. The first treats governance as a single latent construct and aggregates the dimensions into a composite index, often by simple averaging or principal-components extraction (Asongu & Odhiambo, 2019; Khan et al., 2022). The second enters one or several WGI dimensions individually, with the aim of identifying dimension-specific effects on outcomes such as growth, investment, environmental performance, or SDG progress (Aust et al., 2020; Hassan et al., 2022; Wang et al., 2023; Lafortune et al., 2024). Both approaches are defensible in principle, but they are not interchangeable, and the implications of choosing one over the other have received less attention than the underlying substantive questions. More broadly, the choice of how to construct development-related composite indicators carries known methodological consequences for cross-country comparison and policy interpretation, particularly when the underlying indicators are intended to support SDG monitoring (Hák et al., 2016).

Multicollinearity and coefficient instability

Multicollinearity is a well-understood feature of regression analysis. When two or more regressors share a high proportion of common variation, ordinary least squares estimators remain unbiased and consistent under the standard assumptions, but the variance of individual coefficients increases, confidence intervals widen, and small changes in the sample or specification can produce large swings in estimated effects (Wooldridge, 2010; Baltagi, 2021). The variance inflation factor (VIF) is a standard diagnostic, with conventional thresholds of 5 or 10 commonly cited as indicating problematic collinearity. When VIFs exceed these thresholds, the partial coefficients of individual regressors should be interpreted with caution, since they reflect not only the underlying relationship between each regressor and the dependent variable but also the way the shared variation is statistically apportioned across collinear regressors. These concerns echo a broader literature on the responsible construction and interpretation of composite indicators in policy-oriented research, which has highlighted how methodological choices can shape substantive narratives (Kuc-Czarnecka et al., 2020).

A particularly subtle consequence of high collinearity is the appearance of suppressor effects, in which a regressor takes a sign opposite to its bivariate correlation with the dependent variable once a strongly correlated covariate is included in the model (Friedman & Wall, 2005). Suppressor effects do not necessarily indicate model misspecification; in some contexts they reveal genuine partial relationships. However, in cross-country research that draws on highly intercorrelated WGI dimensions, the line between substantive interpretation and statistical artefact is hard to draw. A coefficient that switches sign relative to the bivariate correlation can be misread

as a strong substantive finding when it largely reflects how shared variance is reallocated among collinear governance measures.

Risk- and compliance-relevant implications

Throughout, we distinguish three senses of risk: measurement risk (the choice of governance operationalisation alters the estimated coefficients), inference risk (collinear specifications yield unstable or sign-flipped estimates), and policy-interpretation risk (such estimates may be read as substantive evidence and used to justify policy conclusions). From a governance, risk, compliance, and sustainability research perspective, these statistical considerations carry concrete implications. First, dimension-specific claims about regulatory quality are particularly susceptible to measurement-induced misinterpretation, because regulatory quality is conceptually close to government effectiveness and rule of law ([Kaufmann et al., 2011](#)). Second, policy-oriented work that relies on these claims may reach conclusions that are sensitive to a seemingly minor methodological decision. Third, the appearance of negative coefficients on regulatory quality—if interpreted at face value—could be used to support counterintuitive policy narratives that better regulation undermines sustainability, which would be at odds with both institutional theory and the broader empirical record ([North, 1990](#); [Acemoglu et al., 2001](#); [Rodrik et al., 2004](#); [Glass & Newig, 2019](#)). For these reasons, the present study treats the choice of measurement strategy as a substantive risk in the production of governance–sustainability evidence and seeks to make this risk visible through a transparent comparison.

Theoretical mechanisms linking regulatory quality to sustainable development

Institutional theory holds that development outcomes depend on the rules that structure economic and social behaviour ([North, 1990](#); [Acemoglu et al., 2001](#); [Rodrik et al., 2004](#)). Regulatory quality occupies a specific position within this framework: it captures the state's ability to formulate and implement sound policies and regulations that permit and promote private-sector development ([Kaufmann et al., 2011](#)). Four mechanisms are commonly invoked to explain why this capability should be associated with stronger Sustainable Development Goals performance.

The first is institutional capacity. Formulating environmental standards, social protection rules, and market regulations requires technical expertise, administrative infrastructure, and coordination across agencies. States with higher regulatory quality possess deeper reserves of this capacity, which allows them to design rules that are internally coherent and matched to local conditions rather than transplanted wholesale ([Glass & Newig, 2019](#)).

The second is regulatory effectiveness. Sustainable development outcomes depend less on the existence of rules than on whether those rules alter behaviour. Effective regulation internalises environmental and social externalities, corrects market failures, and creates predictable incentives for long-horizon investment. Where regulatory effectiveness is low, formally adopted standards may generate compliance on paper without measurable improvement in emissions, health, or education outcomes ([Hák et al., 2016](#)).

The third is implementation quality. The gap between adopted policy and realised outcome is mediated by enforcement, monitoring, and administrative follow-through. Higher regulatory quality is associated with more consistent enforcement and lower discretionary variation in how rules are applied, which reduces the uncertainty that deters sustainability-oriented investment ([Biermann et al., 2022](#)).

The fourth is governance efficiency. Regulation imposes compliance costs; efficient regulatory systems achieve their objectives at lower cost to firms and citizens, freeing resources for productive and sustainability-enhancing uses. Inefficient regulation, by contrast, can divert resources into compliance activity that generates little social return.

These four mechanisms motivate an expectation of a positive association between regulatory quality and SDG performance. Critically for the present study, however, none of them is unique to regulatory quality. Institutional capacity, implementation quality, and governance efficiency are equally central to the theoretical content of government effectiveness and rule of law. The conceptual overlap among the WGI dimensions is therefore not incidental but intrinsic: the dimensions are alternative observable manifestations of a partly shared underlying institutional capability. This conceptual overlap is the theoretical counterpart of the statistical collinearity documented below, and it provides the reason to expect that attempts to isolate a dimension-specific effect of regulatory quality will be empirically fragile.

Hypotheses Development

Following [Wooldridge \(2010\)](#), [Baltagi \(2021\)](#), and [Friedman and Wall \(2005\)](#), the study advances three hypotheses about the empirical comparison between the two measurement strategies. These hypotheses concern observable statistical patterns rather than causal mechanisms, consistent with the methodological focus of the study.

- H1. The four WGI dimensions are highly intercorrelated, such that their pairwise correlations are sufficient to produce variance inflation factors above conventional thresholds when the dimensions are entered jointly in a regression model.
- H2. Under Strategy A, the composite governance index is positively associated with SDG performance, consistent with prior cross-country evidence on the governance–sustainability relationship.
- H3. Under Strategy B, the estimated coefficient on Regulatory Quality is statistically unstable—and may exhibit sign reversal—relative to its counterpart in Strategy A, reflecting the redistribution of shared variance across highly collinear WGI dimensions rather than a substantive negative effect of regulatory quality on SDG performance.

RESEARCH METHOD

Data and sample

The study uses a country-year panel covering the 2016–2023 SDG-era window, comprising 1,161 observations across 160 countries. The panel is unbalanced, with an average of approximately 7.3 observations per country. Governance indicators are drawn from the [World Bank \(2024\)](#) Worldwide Governance Indicators, the SDG composite score is taken from [Sachs et al. \(2024\)](#), and the control variables (carbon emissions per capita, Internet users as a share of population, and FDI inflows) are drawn from the World Development Indicators. The data were arranged in country-year format and declared as a panel in Stata using country and year identifiers. The panel was constructed from the raw 2016–2023 country-year file by listwise deletion: an observation enters the estimation sample only if it has non-missing values on the SDG score, all four WGI dimensions, and all controls. Country-year records lacking a valid country identifier were dropped. This yields a complete-case unbalanced panel of 1,161 observations across 160 countries (mean 7.3 years per country); no imputation was used.

Variable measurement

Table 1 summarizes the variables used in the analysis. The dependent variable is the composite SDG score from [Sachs et al. \(2024\)](#), scored on a 0–100 scale. The two measurement strategies for governance are implemented as alternative regressor configurations using the same underlying WGI data: Strategy A enters the row mean of the four dimensions, while Strategy B enters the four dimensions separately. The use of an unweighted row mean follows standard practice in the construction of composite governance indicators, although alternative weighting

and aggregation schemes are known to affect the resulting index's properties and sensitivity (Saisana et al., 2005). Three control variables are included: the natural logarithm of carbon emissions per capita to address skewness in the raw distribution; the share of individuals using the Internet to capture digital access; and FDI inflows winsorized at the 1st and 99th percentiles to mitigate the influence of extreme values; and the natural logarithm of GDP per capita (constant 2015 US\$) to control for the level of economic development. Secondary-school enrolment (% gross) is added as a further control in robustness checks. Although the Worldwide Governance Indicators comprise six dimensions, the present comparison focuses on the four that share the strongest common variation and are most directly implicated in the collinearity that motivates this study; Political Stability and Control of Corruption are excluded to keep the disaggregated specification focused on the dimensions of interest, and their inclusion is left to further work. Year dummies for 2016 through 2023 absorb common time shocks. Year dummies for 2016 through 2023 absorb common time shocks. The selection of variables follows the empirical specification adopted in related cross-country governance–SDG work and is depicted in the research framework (Figure 1).

Table 1. Variable definitions and measurements

Variable	Role	Measurement	Source
SDG score	Dependent variable	Composite sustainable development score, 0–100 scale; higher values indicate stronger SDG performance	Sachs et al. (2024)
Composite governance index	Strategy regressor	A Row mean of Voice & Accountability, Government Effectiveness, Rule of Law, and Regulatory Quality	World Bank (2024) WGI
Voice & Accountability	Strategy regressor	B WGI Voice and Accountability dimension	World Bank (2024) WGI
Government Effectiveness	Strategy regressor	B WGI Government Effectiveness dimension	World Bank (2024) WGI
Rule of Law	Strategy regressor	B WGI Rule of Law dimension	World Bank (2024) WGI
Regulatory Quality	Strategy regressor	B WGI Regulatory Quality dimension	World Bank (2024) WGI
ln_CO2	Control variable	Natural logarithm of carbon emissions per capita	World Development Indicators
Internet users (%)	Control variable	Individuals using the Internet as a percentage of population	World Development Indicators
FDI_w	Control variable	FDI inflows winsorized at 1st and 99th percentiles	World Development Indicators
Year dummies	Time controls	Indicators for years 2016 to 2023	Constructed
ln GDP per capita	Control variable	Natural log of GDP per capita (constant 2015 US\$)	World Bank WDI
Education	Control (robustness)	School enrolment, secondary (% gross)	World Bank WDI

Source: Compiled by author

Research framework

Figure 1 visualizes the research framework. The four WGI dimensions are the common source of governance information, but they enter the empirical model through one of two alternative measurement strategies. Strategy A produces a single composite regressor, which is expected to yield a stable estimate of the governance–SDG association in line with prior literature. Strategy B enters the four dimensions individually, creating the conditions for multicollinearity to redistribute shared variance and to induce sign reversal in highly collinear regressors. Control variables enter both specifications identically. The dashed risk box highlights the methodological concern that motivates the analysis.

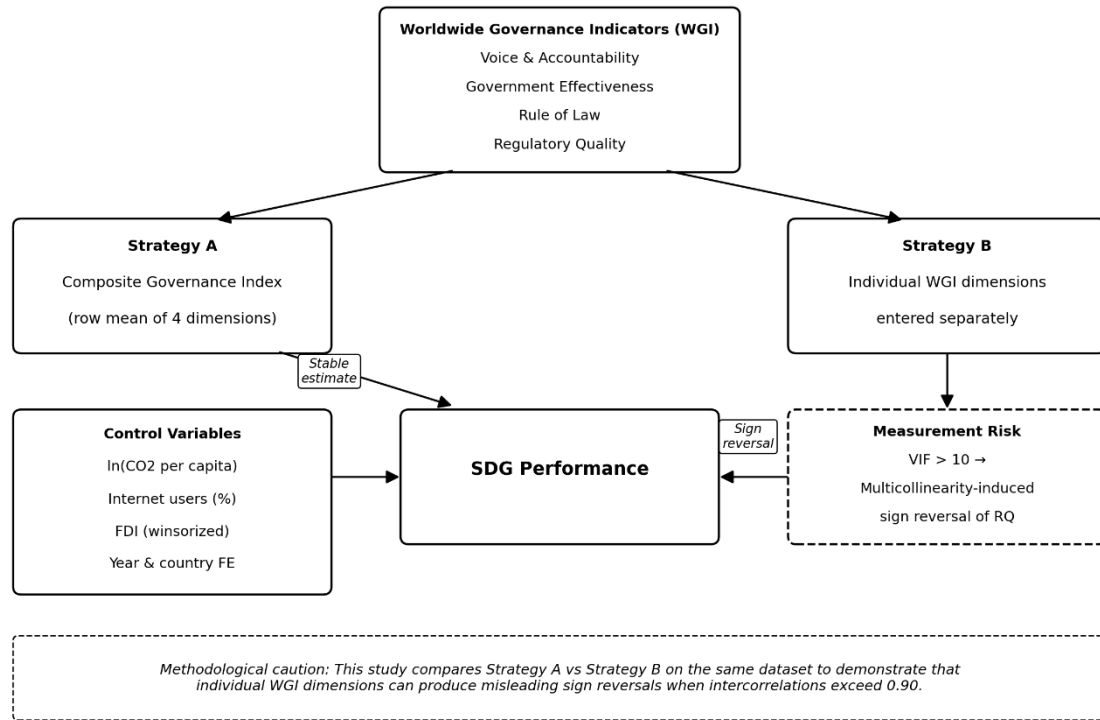


Figure 1. Research Framework

Model specification

Two main specifications are estimated on the same sample. Specification A regresses the SDG score on the composite governance index, control variables, year dummies, and country fixed effects:

$$SDG_{it} = \alpha + \beta_A \cdot Governance_{it} + \gamma' X_{it} + \theta_t + \mu_i + \varepsilon_{it} \quad (1)$$

Specification B replaces the composite index with the four individual WGI dimensions:

$$SDG_{it} = \alpha + \beta_1 \cdot VA_{it} + \beta_2 \cdot GE_{it} + \beta_3 \cdot RL_{it} + \beta_4 \cdot RQ_{it} + \gamma' X_{it} + \theta_t + \mu_i + \varepsilon_{it} \quad (2)$$

where i indexes countries and t indexes years; VA, GE, RL, and RQ denote Voice and Accountability, Government Effectiveness, Rule of Law, and Regulatory Quality, respectively; X_{it} is the vector of controls (ln_CO2, Internet users, and FDI_w, and ln GDP per capita); θ_t denotes year fixed effects; μ_i captures unobserved country-specific effects; and ε_{it} is the idiosyncratic error term.

Each specification is estimated by pooled OLS with heteroskedasticity-robust standard errors, by random-effects GLS, and by fixed-effects regression with cluster-robust standard errors at the country level (Arellano, 1987). Variance inflation factors are computed for each pooled OLS specification. Hausman (1978) tests are used to choose between fixed and random effects. The Wooldridge test for first-order autocorrelation and the Pesaran (2015) CD test for cross-sectional dependence are also reported. To examine whether the sign reversal observed under Strategy B is concentrated in particular regions of the SDG distribution, the analysis additionally reports simultaneous quantile regression estimates at the 0.10, 0.25, 0.50, 0.75, and 0.90 quantiles using 200 bootstrap replications, with the Strategy B configuration of regressors (Koenker, 2005). Throughout, the empirical aim is comparison rather than identification, and the estimates are interpreted as associations rather than causal effects.

FINDINGS AND DISCUSSION

Results

Descriptive statistics

Table 2 summarizes the empirical sample at the variable level. The dependent variable, the composite SDG score, has a mean of 65.75 and a standard deviation of 11.00, with country-year values ranging from 26.10 to 86.80. This range reflects the substantial cross-country heterogeneity in aggregate sustainable development performance during the 2016–2023 window. The four Worldwide Governance Indicators dimensions and their composite are centred close to zero with standard deviations near unity, consistent with the standardized WGI metric reported by Kaufmann et al. (2011). The pairwise similarity in mean and standard deviation across the four dimensions—each with means between -0.038 and -0.026 and standard deviations between 0.931 and 0.996—is itself an early indication of the strong common variation that motivates the methodological comparison developed in subsequent sections. CO2 emissions per capita are highly skewed, with values ranging from approximately zero to over 81 metric tonnes, which justifies the use of the natural logarithm ($\ln_CO2$) in the regression specifications. Internet users as a share of population vary from 1.18 to 100.00 percent, reflecting the wide gap in digital access across countries during the observed period. Foreign direct investment inflows display particularly extreme dispersion before winsorization, ranging from -440 to $+452$ percent of GDP; after winsorization at the 1st and 99th percentiles, FDI_w has a more reasonable range of -10.51 to 103.34 , mitigating the influence of extreme observations on parameter estimates. All statistics are computed on the estimation sample ($N = 1,161$); secondary-school enrolment, used only in robustness checks, is available for 879 observations.

Table 2. Descriptive statistics

Variable	Obs.	Mean	Std. dev.	Min.	Max.
SDG Index	1,161	65.75	11.00	26.10	86.80
Composite governance	1,161	0.01	0.89	-2.06	1.86
Voice & Accountability	1,161	-0.01	0.94	-2.16	1.78
Government Effectiveness	1,161	0.01	0.95	-2.17	2.32
Rule of Law	1,161	-0.03	0.95	-2.31	2.03
Regulatory Quality	1,161	0.05	0.93	-2.00	2.31
CO2 per capita	1,161	4.75	6.09	0.03	48.16
$\ln_CO2$	1,161	0.78	1.43	-3.44	3.87
Internet users (%)	1,161	62.23	27.90	2.20	100.00
FDI (raw, % GDP)	1,161	5.07	33.57	-440.13	452.22
FDI_w (% GDP)	1,161	4.46	11.85	-10.51	103.34

GDP per capita (2015 US\$)	1,161	15,057	20,343	258	110,873
ln GDP per capita	1,161	8.72	1.43	5.55	11.62
Education (sec. enrol. %)	879	90.33	26.49	13.95	160.69

Source: Calculation's author

Correlations and multicollinearity diagnostics

Table 3. Pairwise correlations among WGI dimensions

	Voice & Acc.	Gov. Eff.	Rule of Law	Reg. Quality
Voice & Accountability	1.000			
Government Effectiveness	0.714	1.000		
Rule of Law	0.798	0.937	1.000	
Regulatory Quality	0.748	0.926	0.917	1.000

Source: Calculation's author

Table 3 documents that the four WGI dimensions are highly intercorrelated, consistent with H1. The correlations among Government Effectiveness, Rule of Law, and Regulatory Quality cluster between 0.92 and 0.94, while Voice and Accountability is somewhat less strongly correlated with the other three. These patterns indicate that the four dimensions share a very large proportion of common variation and that any joint specification will need to apportion this shared variation across highly collinear regressors.

Table 4 reports variance inflation factors from the pooled OLS versions of Specifications A and B. Under Strategy A, the composite governance index itself is well within the conventional tolerance (VIF = 4.24); the elevated values in this specification are confined to the control variables (notably ln GDP per capita and ln CO2), reflecting collinearity between governance and the controls rather than within the governance measure. Under Strategy B, by contrast, the governance dimensions themselves are severely collinear: Government Effectiveness, Rule of Law, and Regulatory Quality reach VIFs of 15.66, 13.99, and 12.97 respectively, and the mean VIF rises to 5.85 (versus 3.82 under Strategy A). These diagnostics indicate that the disaggregated configuration operates well above conventional collinearity tolerances and that coefficients on individual WGI dimensions in this configuration should be interpreted with substantial caution.

Table 4. Variance inflation factors

Regressor	VIF (Strategy A)	VIF (Strategy B)
Composite governance index	4.24	–
Voice & Accountability	–	2.88
Government Effectiveness	–	15.66
Rule of Law	–	13.99
Regulatory Quality	–	12.97
ln_CO2	6.68	7.02
Internet users (%)	7.91	8.03
FDI_w	1.02	1.03
ln GDP per capita	12.07	12.12
Year dummies (avg)	≈ 2.0	≈ 2.0
Mean VIF	3.82	5.85

Note. "–" indicates the regressor does not enter that specification (Strategy A uses the composite governance index; Strategy B enters the four WGI dimensions separately). VIFs are computed from the corresponding pooled-OLS model.

Source: Calculation's author

Strategy A versus Strategy B: main panel comparison

Table 5 reports the central empirical comparison. Under Strategy A, the composite governance index enters with a coefficient of 4.20 in the pooled OLS model, statistically significant at conventional levels and consistent with prior cross-country evidence that better governance is associated with stronger SDG performance. Under Strategy B, however, the picture changes. Voice and Accountability and Government Effectiveness enter with positive and significant coefficients, while Rule of Law (-2.09 , $p < 0.01$) and Regulatory Quality (-1.18 , $p < 0.05$) enter with negative coefficients. The aggregate explanatory power of the model is virtually identical across the two strategies (overall $R^2 = 0.80$ versus 0.82), which is expected: the four WGI dimensions and their composite carry essentially the same information but distribute it differently across regressors. The Hausman test rejects the random-effects assumption ($\chi^2(14) = 60.17$, $p < 0.001$), supporting fixed effects for inferential purposes. Crucially, the disaggregated coefficients are not robust to estimator choice: under fixed effects, Rule of Law turns positive and significant ($+3.54$, $p < 0.01$) while Regulatory Quality becomes statistically insignificant, so the negative pooled-OLS signs do not survive the within-country specification preferred by the Hausman test.

Table 5. *Strategy A versus Strategy B: pooled OLS, fixed-effects, and random-effects estimates*

Variable	A: POLS	A: FE	A: RE	B: POLS	B: FE	B: RE
Composite governance	4.195*** (0.361)	0.386 (1.119)	3.174*** (0.719)	–	–	–
Voice & Accountability	–	–	–	2.915*** (0.307)	–2.063 (1.339)	0.720 (0.824)
Government Effectiveness	–	–	–	4.577*** (0.628)	0.566 (0.720)	1.061 (0.698)
Rule of Law	–	–	–	–2.089*** (0.501)	3.538*** (1.080)	2.156** (0.856)
Regulatory Quality	–	–	–	–1.183** (0.507)	–1.397 (0.962)	–0.600 (0.897)
ln_CO2	0.795*** (0.276)	5.426*** (1.144)	2.974*** (0.580)	0.783*** (0.290)	5.227*** (1.050)	2.925*** (0.570)
Internet users (%)	0.237*** (0.014)	0.048** (0.021)	0.075*** (0.021)	0.240*** (0.014)	0.042** (0.021)	0.076*** (0.021)
FDI_w	–0.018** (0.009)	–0.012 (0.009)	–0.010 (0.008)	–0.018* (0.009)	–0.014 (0.009)	–0.011 (0.009)
ln GDP per capita	–0.778** (0.346)	–0.398 (1.295)	0.522 (0.668)	–0.638* (0.337)	–0.854 (1.235)	0.447 (0.668)
Year effects	Yes	Yes	Yes	Yes	Yes	Yes
Country effects	–	Yes (FE)	Random	–	Yes (FE)	Random
Observations	1,161	1,161	1,161	1,161	1,161	1,161
Countries	160	160	160	160	160	160
R ² (overall)	0.803	0.669	0.774	0.817	0.613	0.771
Hausman χ^2 (p)	–	–	–	60.17 (0.000)	–	–

Note. Robust standard errors in parentheses (heteroskedasticity-robust for POLS; clustered at the country level for FE and RE). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Source: Calculation's author

Panel diagnostics support the estimation choices. The Wooldridge test rejects the null of no first-order autocorrelation ($F(1, 151) = 313.08$, $p < 0.001$), so all panel models use standard errors clustered at the country level. The Pesaran (2015) CD test does not reject the null of weak cross-sectional dependence ($CD = -1.51$, $p = 0.131$), giving no evidence that cross-country error correlation distorts the estimates.

Where in the SDG distribution does the sign reversal occur?

Table 6 reports simultaneous quantile regression results under the full Strategy B configuration, with the four WGI dimensions entered separately and year effects included. The coefficient on Regulatory Quality is statistically insignificant at every quantile except the 90th, where it is only marginally significant (-1.11 , $p = 0.06$); at the 10th and 25th quantiles it is small and insignificant (-1.41 , $p = 0.29$; -0.81 , $p = 0.36$). Rule of Law is negative and significant only at the 25th and 50th quantiles and is insignificant elsewhere. Rather than revealing a structured concentration of the negative association at particular quantiles, the estimates show that the coefficients on the individual governance dimensions are unstable in sign and significance across the SDG distribution. Two features of this pattern deserve emphasis. First, the composite specification estimated on the same sample and at the same quantiles yields a positive and statistically significant coefficient at every quantile ($p < .001$ throughout), whereas no individual governance dimension retains a stable sign and significance profile across the distribution. Since the underlying data and the estimator are identical, the difference is attributable to the measurement configuration rather than to quantile estimation as such. Second, the significance of the individual dimensions varies substantially across quantiles while their joint contribution remains comparatively stable, which is the pattern expected when highly correlated regressors share explanatory variance rather than carry independent information.

Table 6. Simultaneous quantile regression estimates under Strategy B

Variable	q10	q25	q50	q75	q90
Voice & Accountability	3.450*** (0.342)	4.032*** (0.498)	2.414*** (0.648)	1.771*** (0.440)	1.445*** (0.553)
Government Effectiveness	4.669*** (0.878)	3.940*** (1.281)	4.204*** (0.903)	2.426*** (0.879)	2.544*** (0.740)
Rule of Law	-1.064 (1.038)	-2.329** (0.970)	-2.200*** (0.629)	0.276 (0.620)	-0.038 (0.793)
Regulatory Quality	-1.408 (1.324)	-0.810 (0.879)	-0.378 (0.655)	-0.505 (0.747)	-1.112* (0.596)
ln_CO2	-0.701* (0.409)	0.211 (0.483)	1.064*** (0.324)	1.555*** (0.319)	1.283*** (0.473)
Internet users (%)	0.235*** (0.024)	0.255*** (0.024)	0.232*** (0.020)	0.214*** (0.020)	0.228*** (0.023)
FDI_w	-0.019 (0.014)	-0.032** (0.016)	-0.003 (0.018)	-0.021* (0.011)	-0.026*** (0.009)
ln GDP per capita	0.206 (0.498)	-0.340 (0.540)	-0.479 (0.510)	-0.796 (0.724)	-0.355 (0.538)
Pseudo R ²	0.602	0.599	0.583	0.564	0.535

Note. Bootstrap standard errors (200 replications) in parentheses. *** $p < .01$, ** $p < .05$, * $p < .10$. $N = 1,161$.

Source: Calculation's author

Diagnostic evidence on shared variance

To provide further support for the interpretation offered above, the sampling behaviour of the Strategy B estimates was examined using a nonparametric bootstrap of the pooled specification with 400 resamples drawn at the country level. This exercise is diagnostic in purpose: it does not alter the estimates or standard errors reported in Tables 5 and 6, but characterises how the coefficients on the individual dimensions behave jointly across resamples. Under collinearity, the estimated coefficients on the affected regressors are not independent, because a resample that apportions more of the shared variance to one dimension apportions less to the others. Negative correlations among the estimates are therefore the expected sampling signature. Table 7 reports these correlations.

The pattern is asymmetric and informative. The estimated coefficient on Government Effectiveness is negatively correlated with those on Rule of Law (−0.655) and Regulatory Quality (−0.467), whereas the estimates for Rule of Law and Regulatory Quality are almost uncorrelated with each other (−0.065). Government Effectiveness thus appears to absorb the shared explanatory variance: resamples in which it takes a larger positive coefficient are systematically those in which the other two dimensions are shifted downward. This corresponds to the configuration reported in Table 5, where Government Effectiveness carries a large positive coefficient while Rule of Law and Regulatory Quality carry negative ones. The diagnostic complements the variance inflation factors in Table 4, which characterise the collinearity of the design matrix but not the joint behaviour of the resulting estimates.

This evidence is consistent with, rather than conclusive of, a multicollinearity explanation. Collinearity diagnostics cannot exclude the possibility that the dimensions capture genuinely divergent institutional properties in particular country groups, and the present analysis does not test that alternative directly. The interpretation is accordingly offered as the most plausible reading of the assembled evidence rather than as a demonstrated mechanism.

Table 7. Bootstrap correlations among estimated coefficients (Strategy B, 400 resamples)

Pair of estimated coefficients	Bootstrap correlation
Government Effectiveness & Rule of Law	−0.655
Government Effectiveness & Regulatory Quality	−0.467
Regulatory Quality & Rule of Law	−0.065

Note. Correlations computed across 400 bootstrap resamples of the Strategy B pooled OLS specification, resampled at the country level (160 clusters). The bootstrap is used as a diagnostic of joint sampling behaviour; reported coefficients and standard errors in Tables 5 and 6 are unchanged.

Source: Calculation's author

Discussion

The central empirical pattern documented in this paper is that the same underlying WGI data, drawn on the same country-year sample and used to explain the same composite SDG outcome, can produce quite different substantive conclusions depending on the chosen measurement strategy. Under Strategy A, the composite governance index is positively and significantly associated with SDG performance, in line with the broader institutional view of development (North, 1990; Acemoglu et al., 2001; Rodrik et al., 2004; Glass & Newig, 2019;

[Lafortune et al., 2024](#)). Under Strategy B, the same underlying information is reapportioned across four highly collinear regressors, and Regulatory Quality and Rule of Law enter with negative coefficients while Government Effectiveness absorbs a large positive coefficient. The variance inflation factors and the quantile pattern together suggest that the negative signs on Regulatory Quality and Rule of Law in Strategy B are most plausibly interpreted as consequences of multicollinearity rather than as evidence that better regulation undermines sustainable development. Consistent with this reading, the negative signs do not survive estimation with fixed effects, the specification preferred by the Hausman test: under fixed effects, Rule of Law enters positively and Regulatory Quality is no longer statistically significant.

This interpretation should be made cautiously. Collinearity diagnostics are necessary but not always sufficient to rule out genuine differences across WGI dimensions. In principle, regulatory quality and government effectiveness could move in partly opposite directions in some country clusters, and the disaggregated specification might capture an underlying institutional tension that the composite index averages away. The present study cannot, with the data at hand, fully separate these alternative readings. What it can document is that the observed sign reversal coincides precisely with the conditions under which multicollinearity is known to redistribute coefficients—high pairwise correlations and elevated VIFs—and that the disaggregated coefficients are unstable across estimators and across the SDG distribution. The most defensible conclusion is therefore that the negative coefficient on regulatory quality in Strategy B should not be treated as a stand-alone substantive finding without further triangulation, such as alternative measurement schemes, instrumental variables, or qualitative case evidence.

The implications for governance, risk, compliance, and sustainability research are practical. First, when WGI dimensions are entered jointly, researchers should report variance inflation factors and pairwise correlations, and should interpret coefficients on individual dimensions with caution when collinearity diagnostics indicate elevated values. Second, where the substantive question concerns the general role of governance quality, a composite index, principal-components factor, or hierarchically structured measure may convey more reliable information than four separately entered dimensions ([OECD/JRC, 2008](#)). Third, where the substantive question genuinely requires dimension-specific claims—for example, the distinct contribution of regulatory quality to compliance and risk outcomes—researchers should consider alternative measurement instruments that draw on more heterogeneous underlying data sources than the WGI alone, or should triangulate WGI-based estimates with implementation, audit, or compliance-rate indicators. Fourth, reviewers and readers can use the diagnostic pattern documented here as a checklist for evaluating dimension-specific claims about regulatory quality.

The findings also speak to the practice of frame-setting in policy-oriented research. A negatively signed coefficient on regulatory quality, presented without collinearity diagnostics, could plausibly be used to support a counterintuitive narrative—for example, that stricter regulation harms sustainability—that is at odds with the broader theoretical and empirical record. The methodological caution advanced here is that such narratives should be supported by more than a single specification with elevated VIFs. From a risk management perspective, the choice of measurement strategy is itself a source of evidence-production risk, and recognizing this risk explicitly is part of responsible governance research.

Two additional robustness checks reinforce this reading. First, replacing the four separate dimensions with a single principal-components governance index — which loads positively and almost equally on all four dimensions (first component explaining 89% of their joint variance) — yields a positive and significant coefficient (1.96, $p < 0.01$), mirroring the composite-index result under Strategy A. Second, dropping the most collinear dimension (Rule of Law) and retaining Government Effectiveness, Regulatory Quality, and Voice and Accountability does not restore a

positive sign on Regulatory Quality (-1.79 , $p < 0.01$); because Government Effectiveness and Regulatory Quality remain correlated at 0.93 , the disaggregated coefficient stays unstable. Together these checks indicate that the negative sign on Regulatory Quality tracks the collinearity structure of the disaggregated specification rather than a substantive relationship.

CONCLUSIONS

This study has compared two measurement strategies for governance in a cross-country panel of 160 countries observed during 2016–2023. The composite governance index produces the conventional positive association between governance and SDG performance. The disaggregated specification produces negative coefficients for Regulatory Quality and Rule of Law, against a backdrop of variance inflation factors above the conventional threshold and pairwise correlations exceeding 0.90 among the affected dimensions. Quantile and fixed-effects estimates show that these negative associations are not robust across the SDG distribution or across estimators. Read together, the patterns are most plausibly interpreted as a measurement artefact rather than as substantive evidence that regulatory quality reduces SDG performance.

Theoretical contribution

The study makes three contributions to governance and sustainability research. First, it specifies the theoretical reason why dimension-specific claims about regulatory quality are fragile: institutional capacity, regulatory effectiveness, implementation quality, and governance efficiency are mechanisms shared across the WGI dimensions rather than proprietary to any one of them, so the dimensions are best understood as alternative manifestations of a partly common institutional capability. The statistical collinearity documented here is therefore the empirical counterpart of a conceptual overlap, not merely a data problem.

Second, it clarifies the level at which the institutional argument is empirically supported. The evidence sustains the proposition that general governance quality is positively associated with sustainable development, but it does not sustain claims about the separate contribution of individual governance dimensions estimated from WGI data alone. This distinction matters for theory building, because it identifies the resolution at which the WGI can and cannot adjudicate institutional arguments.

Third, it contributes to the measurement literature on composite indicators by demonstrating that specification dependence in this setting is strong enough to reverse the sign of a theoretically central variable. This extends the general caution about composite indicator construction (Saisana et al., 2005; OECD/JRC, 2008; Kuc-Czarnecka et al., 2020) to the specific case of governance–sustainability inference.

Practical and policy implications

For researchers, the practical takeaway is that the choice between composite and disaggregated WGI specifications should be treated as a consequential measurement decision rather than a neutral implementation detail. Where the research question concerns general governance quality, a composite index or a factor-based measure is preferable. Where individual dimensions are entered jointly, variance inflation factors and pairwise correlations should be reported as a matter of routine, and coefficient stability should be verified across estimators before any dimension-specific coefficient is interpreted substantively.

For journal editors and reviewers, the diagnostic pattern documented here offers a practical screening checklist for evaluating manuscripts that advance dimension-specific claims about regulatory quality: whether collinearity diagnostics are reported, whether results are shown across alternative estimators, and whether the sign of the reported coefficient is stable.

For policymakers and analysts who use governance indicators to inform regulatory reform, the implication is a caution against acting on a negatively signed coefficient for regulatory quality obtained from a single disaggregated specification. Such a coefficient does not provide a defensible basis for concluding that regulatory tightening impedes sustainable development. Where dimension-specific policy guidance is genuinely required, WGI-based estimates should be triangulated with implementation, audit, or compliance-rate indicators that draw on more heterogeneous underlying sources.

LIMITATION & FURTHER RESEARCH

The present study is a methodological note based on a single cross-country panel and is best understood as one piece of evidence within a broader research conversation, rather than as a definitive treatment of governance measurement. Several limitations are worth acknowledging openly, and each in turn motivates promising avenues for further work.

First, the analysis focuses on four WGI dimensions and their composite. Future research could extend the comparison to alternative governance measurement schemes, including principal-components factor scores, hierarchical Bayesian latent-variable models, and structurally specified second-order constructs. Such extensions would help characterize the conditions under which dimension-specific WGI claims can be sustained on statistical grounds.

Second, the empirical design is observational and associational, and is not intended to identify causal effects. Future studies could complement the present analysis with instrumental-variable approaches, natural-experiment designs around major regulatory reforms, or system generalized method of moments specifications, all of which would help disentangle whether the differences across WGI dimensions reflect substantive institutional dynamics or statistical artefacts.

Third, the sample window of 2016–2023 is relatively short. Longer panels would allow more precise within-country estimation of governance effects and would provide a stronger test of whether the sign reversal documented here persists across business cycles, regional sub-samples, and structural breaks. Future research could re-estimate the comparison over extended SDG horizons as more data become available.

Fourth, the analysis now controls for GDP per capita (constant 2015 US\$) in the main specification and for secondary-school enrolment in robustness, and the multicollinearity pattern documented here is robust to their inclusion. Other development-relevant covariates—such as trade openness, demographic structure, and fiscal capacity—are not included and remain avenues for further work; future studies could also examine whether interactions between governance and income or trade openness yield more stable dimension-specific estimates ([Asongu & Odhiambo, 2019](#); [Aust et al., 2020](#)).

Fifth, the SDG outcome is operationalized as a composite score. Future research could disaggregate the dependent variable into thematic SDG pillars or specific indicators, examining whether the measurement risk identified here varies across environmental, social, and economic SDG domains.

Finally, beyond methodological refinement, there are several promising substantive directions. Subsequent work could combine the WGI with direct compliance and enforcement measures—such as audit indicators, environmental violation records, or implementation-rate metrics—to triangulate dimension-specific claims about regulatory quality. Studies could also examine whether the composite-versus-disaggregated divergence is more or less pronounced in specific policy domains, including environmental regulation, financial regulation, and digital governance. These extensions would build on the methodological caution offered in this paper and contribute to a more robust evidence base for governance, risk, compliance, and sustainability

research.

REFERENCES

- Acemoglu, D., Johnson, S., & Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review*, 91(5), 1369–1401. <https://doi.org/10.1257/aer.91.5.1369>
- Arellano, M. (1987). Computing robust standard errors for within-groups estimators. *Oxford Bulletin of Economics and Statistics*, 49(4), 431–434. <https://doi.org/10.1111/j.1468-0084.1987.mp49004006.x>
- Asongu, S. A., & Odhiambo, N. M. (2019). Governance, capital flight and industrialisation in Africa. *Journal of Economic Structures*, 8, Article 36. <https://doi.org/10.1186/s40008-019-0170-2>
- Aust, V., Morais, A. I., & Pinto, I. (2020). How does foreign direct investment contribute to Sustainable Development Goals? Evidence from African countries. *Journal of Cleaner Production*, 245, Article 118823. <https://doi.org/10.1016/j.jclepro.2019.118823>
- Baltagi, B. H. (2021). *Econometric analysis of panel data* (6th ed.). Springer. <https://doi.org/10.1007/978-3-030-53953-5>
- Biermann, F., Hickmann, T., Sénit, C.-A., Beisheim, M., Bernstein, S., Chasek, P., Grob, L., Kim, R. E., Kotzé, L. J., Nilsson, M., Ordóñez Llanos, A., Okereke, C., Pradhan, P., Raven, R., Sun, Y., Vijge, M. J., van Vuuren, D., & Wicke, B. (2022). Scientific evidence on the political impact of the Sustainable Development Goals. *Nature Sustainability*, 5(9), 795–800. <https://doi.org/10.1038/s41893-022-00909-5>
- Friedman, L., & Wall, M. (2005). Graphical views of suppression and multicollinearity in multiple linear regression. *The American Statistician*, 59(2), 127–136. <https://doi.org/10.1198/000313005X41337>
- Glass, L.-M., & Newig, J. (2019). Governance for achieving the Sustainable Development Goals: How important are participation, policy coherence, reflexivity, adaptation, and democratic institutions? *Earth System Governance*, 2, Article 100031. <https://doi.org/10.1016/j.esg.2019.100031>
- Hák, T., Janoušková, S., & Moldan, B. (2016). Sustainable Development Goals: A need for relevant indicators. *Ecological Indicators*, 60, 565–573. <https://doi.org/10.1016/j.ecolind.2015.08.003>
- Hassan, S. T., Batool, B., Sadiq, M., & Zhu, B. (2022). How do green energy investment, economic policy uncertainty, and natural resources affect greenhouse gas emissions? A Markov-switching equilibrium approach. *Environmental Impact Assessment Review*, 97, Article 106887. <https://doi.org/10.1016/j.eiar.2022.106887>
- Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, 46(6), 1251–1271. <https://doi.org/10.2307/1913827>
- Kaufmann, D., Kraay, A., & Mastruzzi, M. (2011). The Worldwide Governance Indicators: Methodology and analytical issues. *Hague Journal on the Rule of Law*, 3(2), 220–246. <https://doi.org/10.1017/S1876404511200046>
- Khan, S. A. R., Ponce, P., Tanveer, M., Aguirre-Padilla, N., Mahmood, H., & Shah, S. A. A. (2022). Technological innovation and environmental taxes toward a carbon-free economy: An empirical study in the context of COP-21. *Journal of Environmental Management*, 298, Article 113418. <https://doi.org/10.1016/j.jenvman.2021.113418>
- Koenker, R. (2005). *Quantile regression*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511754098>
- Kuc-Czarnecka, M., Lo Piano, S., & Saltelli, A. (2020). Quantitative storytelling in the making of a composite indicator. *Social Indicators Research*, 149, 775–802.

<https://doi.org/10.1007/s11205-020-02276-0>

- Lafortune, G., Fuller, G., Bermont-Diaz, L., Klope-Lesch, A., Koundouri, P., & Riccaboni, A. (2024). *European sustainable development report 2023/24: The European elections, Europe's future and the Sustainable Development Goals*. Sustainable Development Solutions Network; SDSN Europe; Dublin University Press. <https://doi.org/10.25546/104407>
- North, D. C. (1990). *Institutions, institutional change and economic performance*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511808678>
- OECD/JRC. (2008). *Handbook on constructing composite indicators: Methodology and user guide*. OECD Publishing. <https://doi.org/10.1787/9789264043466-en>
- Pesaran, M. H. (2015). Testing weak cross-sectional dependence in large panels. *Econometric Reviews*, 34(6–10), 1089–1117. <https://doi.org/10.1080/07474938.2014.956623>
- Rodrik, D., Subramanian, A., & Trebbi, F. (2004). Institutions rule: The primacy of institutions over geography and integration in economic development. *Journal of Economic Growth*, 9(2), 131–165. <https://doi.org/10.1023/B:JOEG.0000031425.72248.85>
- Sachs, J. D., Lafortune, G., & Fuller, G. (2024). *The SDGs and the UN Summit of the Future: Sustainable development report 2024*. Sustainable Development Solutions Network; Dublin University Press. <https://doi.org/10.25546/108572>
- Sachs, J. D., Lafortune, G., Kroll, C., Fuller, G., & Woelm, F. (2022). *From crisis to sustainable development: The SDGs as roadmap to 2030 and beyond. Sustainable development report 2022*. Cambridge University Press. <https://doi.org/10.1017/9781009210058>
- Saisana, M., Saltelli, A., & Tarantola, S. (2005). Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168(2), 307–323. <https://doi.org/10.1111/j.1467-985X.2005.00350.x>
- Wang, Q., Zhang, F., & Li, R. (2023). Free trade and carbon emissions revisited: The asymmetric impacts of trade diversification and trade openness. *Sustainable Development*, 31(1), 219–239. <https://doi.org/10.1002/sd.2387>
- Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). MIT Press.
- World Bank. (2024). *Worldwide Governance Indicators* [Data set]. World Bank Group. <https://www.worldbank.org/en/publication/worldwide-governance-indicators>