Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 4 No. 3 (March) (2026)

Governance Components and the Tax–Growth Association in Afghanistan (1996–2024): Matched Interaction Contrasts and Uncertainty

Dr. Wafiullah Kareemzada
PhD in Finance, Institute of Management Studies, University of Peshawar, Pakistan
Email: wafikareemzada@uop.edu.pk
ORCID: https://orcid.org/0009-0009-7397-2509

ABSTRACT
This article directly examines whether three governance dimensions show distinguishable tax–growth interactions in Afghanistan's annual 1996–2024 archive (N = 29). Archived single-component OLS regressions are independently reproduced: tax × Government Effectiveness = 4.6158 (p = .221), tax × Rule of Law = 9.7837 (p = .371) and tax × Control of Corruption = −0.6010 (p = .940). A new matched, jointly estimated specification standardizes all three components to their observed one-standard-deviation scales while preserving the same years, FDI, trade and population controls. A conventional test for equal standardized interactions yields F(2,18) = 0.232 (p = .795); an HC3 test yields p = .764. A joint test that all three interaction coefficients equal zero also fails to reject (p = .784 conventional; p = .750 HC3). High regressor dependence and very small residual degrees of freedom limit interpretability. The dynamic ARDL(1,0,0,1,0) from 1997–2020 (N = 24) is contextual, not a component moderation test. The new direct contrasts do not support ranking governance components or claiming fiscal-growth policy effects; identification limits, WGI-vintage uncertainty, and overlap with previously published research are acknowledged.

Keywords: Afghanistan; governance components; government effectiveness; rule of law; control of corruption; taxation; interaction models; fragile states

1. Introduction

Domestic tax revenue potentially finances public infrastructure, basic services, and administrative functions that improve private-sector productivity. Whether revenue mobilisation accompanies higher economic growth, however, depends on the economic costs of taxation, the allocation of expenditure, and institutions that implement public decisions. These conditions are particularly important in conflict-affected economies, where administrative continuity, tax compliance, external assistance, and national accounts measurement can change abruptly. Afghanistan's political and economic experience between 1996 and 2024 makes this relationship important but difficult to estimate reliably. A national time series offers valuable historical coverage while providing few observations with which to distinguish correlated institutional mechanisms (Akitoby et al., 2020; Besley & Persson, 2009).

Governance is not one indivisible policy instrument. Government Effectiveness (GE) addresses perceptions of the quality of public services and policy implementation; Rule of Law (RL) concerns the predictability and enforcement of formal rules; and Control of Corruption (CC) concerns the perceived use of public authority for private gain. Each may be relevant to fiscal administration, yet the empirical claim that one component provides a greater growth payoff than another is stronger than the claim that any individual

ISSN Online: 3007-3154
ISSN Print: 3007-3146

component is associated with the tax–growth slope. Both claims require distinct statistical tests. The World Bank's Worldwide Governance Indicators (WGI) provide useful proxies, but the indicators are perception-based and are not direct audits of Afghan revenue offices or expenditure programs (Kaufmann et al., 2011; World Bank, 2026).

This article examines the evidence for three separate tax-by-governance interaction coefficients recovered from the corrected research archive. Its focus is the precision and comparability of the conditional tax–growth associations, not the construction of a governance league table. The reported interactions are positive for GE and RL and slightly negative for CC, but all are statistically imprecise. The question is therefore what, if anything, these results establish about institutional conditioning in Afghanistan. This is especially relevant because point estimates from different regressions cannot be subtracted as though they came from a matched joint model with a known cross-equation covariance matrix (Brambor et al., 2006).

The archive includes a separately selected ARDL(1,0,0,1,0) with 1997–2020 effective observations (N = 24). It contains neither the governance components nor their interactions, and is presented as previously reported context rather than confirmatory component evidence. The new joint OLS contrasts are the article's principal analysis (Kripfganz & Schneider, 2023; Pesaran et al., 2001).

The paper has a specific relationship to the published study by Kareemzada et al. (2026), which already analysed Afghanistan's taxation, institutional quality, political transitions, and ARDL–ECM evidence using an overlapping source archive. The present contribution is narrower: a component-specific examination of uncertainty, coefficient comparability, and measurement vintage. It does not claim an independent dataset, a new causal design, or a formally established component ordering. The empirical contribution is restricted to new same-sample interaction contrasts and their reported uncertainty.

The study asks four research questions: (RQ1) What do archived individual tax-by-governance interactions establish? (RQ2) Are those results robust to HC3 covariance estimation? (RQ3) Can standardized interactions be compared within a matched, jointly estimated model? (RQ4) Does a formal equality restriction distinguish Government Effectiveness, Rule of Law, and Control of Corruption? The dynamic ARDL is contextual only. The article explicitly separates archived estimates from the independently executed same-sample, standardized comparison.

2. Literature Review and Conceptual Framework

2.1 Taxation, public inputs, and fiscal capacity

Economic theory provides reasons for either a positive or a negative observed tax–growth association. In Barro's (1990) model, some government services raise the productivity of private capital, yet the taxation needed to finance them also imposes costs. Consequently, more taxation does not mechanically imply more growth; the revenue-financed composition of spending and the tax burden jointly matter. North (1990) explains how enforceable rules shape investment incentives and transaction costs. Together these arguments suggest that the quality of implementing institutions could condition the association between a country's recorded revenue ratio and its subsequent economic performance, without predetermining the empirical sign.

State-capacity theory extends the analysis beyond tax rates. Besley and Persson (2009) model legal and fiscal capacity as complementary investments shaped by political incentives and historical constraints. Fiscal collection depends on the ability to identify liabilities, monitor transactions, enforce rules, and process disputes. Moore (2004) and Prichard (2015) further discuss how reliance on domestic taxpayers may create accountability bargains, although such bargains vary by political context. These

ISSN Online: 3007-3154
ISSN Print: 3007-3146

mechanisms are plausible in Afghanistan but are not directly identified by an annual GDP-growth regression. Distinguishing the conceptual channel from its observed proxy is essential: WGI component scores do not measure tax administration productivity or citizens' individual compliance decisions.

Evidence on fragile and conflict-affected states shows that domestic mobilisation is associated with considerable institutional and administrative difficulties. Akitoby et al. (2020), examining revenue mobilisation in these settings, emphasise tax policy and administration reform rather than an automatic revenue-to-growth multiplier. Afghanistan presents additional challenges because aid, trade dependence, political upheaval, and changes in economic reporting can influence both recorded taxes and GDP growth. With only 29 annual observations, an interaction term may also respond disproportionately to a small number of unusual years. The literature therefore motivates careful controls, transparent observational definitions, and appropriate uncertainty rather than simple causal claims.

The revenue measure is itself a ratio, not a randomized change in the tax rate. Higher tax-to-GDP collections may accompany economic formalisation, stronger customs operations, domestic demand, or changes in the GDP denominator. The theoretical public-input channel also depends on whether revenue is actually spent on services that raise productive capacity; the annual revenue ratio does not observe expenditure composition. In fragile settings, off-budget assistance or donor-financed public goods may weaken the correspondence between domestic collections and services delivered. The analysis therefore interprets a positive tax coefficient as an association conditional on available macroeconomic controls, not as a marginal productivity estimate of public revenue. This distinction also prevents the literature review from importing theoretical effects about public spending into tests that contain no public-spending variable (Barro, 1990; Akitoby et al., 2020).

2.2 Three governance mechanisms

Government Effectiveness represents perceptions of the quality of public services, policy formulation, implementation, and civil-service credibility. Its hypothesised fiscal role is operational: reliable administration can reduce arbitrary assessments and processing delays, while effective expenditure management can improve the use of collected revenue. The corresponding empirical expectation is a positive TR × GE coefficient. Nevertheless, the WGI estimate cannot indicate how much of a tax payment becomes productive public capital, and an interaction with GE alone does not test the intervening operational activities.

Rule of Law reflects perceptions concerning confidence in legal rules, contract enforcement, property rights, and related institutions. The expected mechanism is predictability: clearer tax rules and appeals, secure transactions, and less discretionary enforcement may support formal investment and a more stable tax base. An interaction with RL asks whether the fitted tax–growth gradient changes with the observed legal-governance proxy. It does not estimate the effect of a particular judicial reform or distinguish legal improvements from correlated political and economic developments (North, 1990; Kaufmann et al., 2011).

Control of Corruption is relevant to leakage, procurement integrity, exemptions, enforcement discretion, and the diversion of public resources. A tax system perceived as less corrupt might support trust and more efficient public expenditure, theoretically strengthening the tax–growth relationship. Yet perceptions of corruption, capacity, and political authority overlap, and integrity improvements may affect growth only when combined with functioning administrative and legal systems. An estimated TR × CC

ISSN Online: 3007-3154
ISSN Print: 3007-3146

interaction is consequently a conditional association, not an observed estimate of money recovered through anti-corruption reform.

The components are conceptually related, not statistically independent. Besley and Persson (2009) imply complementarity between fiscal and legal capabilities, while WGI source judgments can load on more than one broad governance dimension. Interpreting separate coefficient magnitudes as directly comparable requires evidence of matched control sets, common support, an appropriate transformation to comparable governance changes, and an inferential procedure for differences. None can be obtained merely by noticing that the absolute value of one coefficient exceeds another. Moreover, a non significant interaction may be consistent with both economically consequential positive and negative conditioning because of a wide confidence interval (Brambor et al., 2006).

An additional conceptual issue is that the WGI components describe broad institutional conditions rather than policy levers that can be adjusted independently. Improvements in legal enforcement may require capable public administration, while anti-corruption enforcement relies on effective legal procedures and public records. Such complementarities can make an isolated component regression difficult to interpret even if its estimate were statistically precise. A meaningful comparison should distinguish a governance score associated with many simultaneous social changes from a discrete action such as implementing electronic filing, creating an independent tax appeals process, or publishing procurement awards. These specific reforms could affect compliance costs and expenditure credibility through different horizons. No reform-specific implementation date, treatment group, or operational outcome appears in the archival equations, so their mechanisms remain theoretical rather than empirically tested. State-capacity theory supports treating them as potential complements instead of automatically competing alternatives (Besley & Persson, 2009; Prichard, 2015).

2.3 Research hypotheses and model alignment

The hypotheses are distinguished by estimation status. H1 concerns the long-run tax coefficient in the previously estimated ARDL and is retained only as contextual evidence. H2–H4 concern whether the tax–growth slope varies with Government Effectiveness, Rule of Law, and Control of Corruption in the separate static interaction models. An additional comparative null tests equality of the three standardized interaction coefficients in one matched specification (Table 5: conventional $p = .795$; HC3 $p = .764$). Failure to reject this restriction neither establishes equivalent effects nor permits a ranking of the governance dimensions.

The conceptual model treats tax revenue as the focal explanatory variable, annual real GDP growth as the measured outcome, and one governance component at a time as a moderator. Foreign direct investment (FDI), trade openness, and population growth enter the described component specifications as controls. Productive service delivery, legal predictability, and reduced leakage are proposed theoretical mechanisms, not estimated mediators. Figure 1 makes this distinction visible. It also avoids drawing a political-period mechanism or threshold arrow not tested by the reported component equations.

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 4 No. 3 (March) (2026)

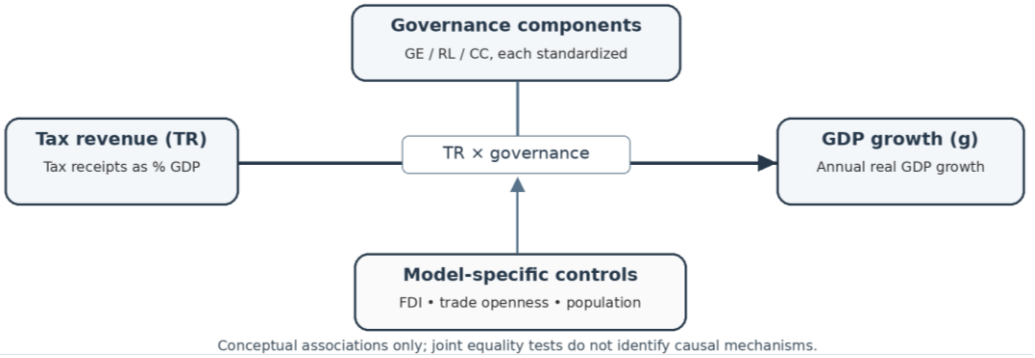


Figure 1: Conceptual model of component-specific governance moderation

3. Data and Econometric Methodology

3.1 Observation windows, variables, and data integrity

The harmonised country-level master archive covers 1996–2024 and contains 29 annual observations. Growth is the annual percentage change in real GDP; tax revenue is expressed as a percentage of GDP. The documented controls are FDI net inflows, trade openness (both expressed relative to GDP), and annual population growth. Candidate international sources include the World Development Indicators (WDI), while national cross-checks were recorded from the National Statistics and Information Authority (NSIA), Da Afghanistan Bank (DAB), and Afghanistan's Ministry of Finance (MoF). The harmonization protocol draws on WDI and Afghan administrative sources; however, the observation-level reconciliation ledger is not available for verification. No missing annual value is reconstructed or replaced in the present article.

The original source protocol distinguishes O (official), E (estimated), P (proxy), and A (administratively adjusted) observations. Such classes matter because an imputed, proxy, or zero-filled series entry cannot automatically be treated as a directly observed fiscal outcome. The master-file descriptive table records minimum values of zero for both FDI and trade; without the observation-level ledger it is impossible to establish whether each zero reflects a measured economic absence or a treatment of unavailable data. Appendix A specifies the required audit trail. Historical Afghan fiscal years may differ from international calendar-year reporting; each reconciliation requires an explicit source date, conversion method, original value, and final harmonised value. The manuscript cannot certify those transformations without the corresponding underlying files.

WGI provides estimates for six conceptual dimensions, but the dissertation's five-item composite averages CC, RL, GE, Voice and Accountability, and Regulatory Quality. Political Stability and Absence of Violence/Terrorism is excluded from this specific composite. The present article does not average the three chosen components into a new index: it estimates each moderator separately. The estimated models were produced from an archived governance series whose exact release and download date are not identified in the supplied article. This is a nontrivial reproducibility limitation. The World Bank's 2025 methodological revision and 2026 release recalculated historical estimates back to 1996; the 2026 data cannot reproduce the earlier Stata coefficients without a genuinely new model estimation (World Bank, 2026).

The effective sample differs by analysis. Descriptive statistics and component OLS specifications use 1996–2024 (N = 29); the controlling ARDL is a restricted 1997–2020 sample (N = 24). These are not substitute windows. The contextual benchmark is the BIC-selected ARDL(1,0,0,1,0) estimated over 1997–2020 (N = 24), without governance regressors. Regime periods are historical context, not causal treatment assignments, and no period-specific moderation parameter is presented as confirmed evidence.

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 4 No. 3 (March) (2026)

Each annual observation represents one country-year, so standard cross-sectional arguments about averaging over many independent units do not apply. Afghanistan's persistent governance scores may be strongly serially correlated, and the number of major political transitions is limited. The interaction regressions can detect a pattern only to the extent that tax revenue varies meaningfully within the available governance range. If nearly all low-quality scores occur during one economic era and nearly all relatively higher scores during another, the product coefficient may partially reflect period-specific shocks. The source article provides no complete influence diagnostics for the three component equations and does not establish that coefficients are insensitive to excluding extraordinary years. These uncertainties are an appropriate topic for the limitations section rather than a justification for choosing favorable regression outputs.

3.2 Separate governance interaction regressions

For each $j \in \{GE, RL, CC\}$, define $G^c_{j,t} = G_{j,t} - \text{mean}(G_j)$ over the 29 observations. We independently reproduce three static OLS models with tax revenue, the component main effect, their product, FDI, trade openness and population growth. The corresponding estimates and HC3 covariance calculations appear in Table 4. For a directly matched comparison, all three components are standardized to observed standard deviations and included together, along with their tax interactions and the same controls, in a joint OLS regression; formal equalities use its within-model covariance matrix (Table 5). Original source-vintage authentication is a separate outstanding requirement.

$$g_t = \alpha_j + \beta_{1j} TR_t + \beta_{2j} G^c_{j,t} + \kappa_j (TR_t \times G^c_{j,t}) + \gamma_j' X_t + u_{j,t} \qquad (1)$$

Here κj is the change in the fitted tax–growth slope associated with a one-unit increase in that centred governance score; it is not the causal effect of improving that institution. Conditional on X and the maintained linear model, Equation 2 defines the fitted tax slope at centred governance state z. With mean centring, $\beta 1j$ denotes the tax slope at the component's sample mean. This constituent coefficient is not a cross-model average treatment effect.

$$m_j(z) = \partial E[g_t \mid TR_t, G^c_{j,t}, X_t] / \partial TR_t = \beta_{1j} + \kappa_j z \qquad (2)$$

For inference about m_j(z), the complete within-model coefficient covariance matrix is required. The conditional-slope variance is $\text{Var}(\beta 1j) + z^2\text{Var}(\kappa j) + 2z\text{Cov}(\beta 1j, \kappa j)$; a confidence interval also requires the relevant degrees of freedom and test reference distribution. The final covariance term cannot be presumed zero. Consequently, no conditional slope bands, region-of-significance estimates, or Johnson–Neyman points are reproduced here. The source does report coefficient-specific confidence intervals for κj, which are retained verbatim in Table 3. A significance region for a conditional slope, were one to be calculated, would not constitute a structural governance threshold (Brambor et al., 2006).

The original component coefficients refer to different governance scales and cannot be ranked by size or separate p values. Table 5 provides an independently executed common-sample joint specification, standardizing GE, RL and CC by their own observed standard deviations and directly testing differences with conventional and HC3 covariance matrices. Its high multicollinearity and 18 residual degrees of freedom prevent a precise component ranking. Distinguishing this estimated comparison from a causal institutional intervention is essential.

3.3 Conditions for direct component comparisons

Tests of each individual interaction against zero do not imply that their coefficients differ. Table 5 uses a jointly estimated, standardized model to test the equalities H_0: $\kappa GE = \kappa RL = \kappa CC$ and the separate joint-zero restriction. Its conventional equality result is $F(2, 18) = 0.232$ ($p = .795$); the HC3 joint result has $p = .764$. None establishes a ranking. These are

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 4 No. 3 (March) (2026)

independently calculated comparisons, not values reconstructed by subtracting archival model coefficients; they have their own covariance matrix and model identification. A limited sample and shared governance variation constrain precision even when the test is properly specified.

The component equations are static contemporaneous associations, not interventions. Revenue may respond to output growth, governance perceptions may respond to macroeconomic shocks, and common denominators can couple the measurement of tax revenue and trade with output. Lagging or instrumenting regressors might be considered in future research, but no new identification strategy is claimed to have been estimated. The annual sample is small relative to the complexity of any simultaneous governance model, and the number of feasible lags or additional interaction terms is limited. Consequently, the methods support constrained inference about conditional associations rather than causal fiscal returns.

Mean-centring changes the reference value of a constituent tax coefficient but, with both main effects included, not the tax-by-governance product coefficient. Standardization changes interaction units and requires a matched variance-covariance calculation. The individual component equations and the new joint specification are therefore reported as separate models. Their signs and confidence intervals cannot be interpreted as effects of an intervention or as evidence that one government function should be privileged over the others.

3.4 Distinct dynamic benchmark and diagnostics

The corrected dynamic baseline is the Bayesian information criterion (BIC)-selected ARDL(1,0,0,1,0) for growth, tax revenue, FDI, trade openness, and population growth. In a compact levels parameterisation it includes lagged growth; contemporaneous tax, FDI, and population growth; and both current and lagged trade openness. As a schematic specification, Equation 3 clarifies timing without implying that the model includes a governance variable or a period interaction. The levels equation uses growth, tax revenue, FDI, trade openness and population growth as defined in Section 3.1.

$$g_t = a + \rho g_{t-1} + b_1 TR_t + b_2 FDI_t + b_3 Trade_t + b_4 Trade_{t-1} + b_5 Pop_t + \varepsilon_t \quad (3)$$

When a levels relation is identified and the normalisation is admissible, the long-run tax coefficient is $b1/(1 - \rho)$, whereas the error-correction adjustment parameter is $\rho - 1$. In the source run, the bounds test is Case 3 with four forcing regressors and N = 24. Its observed F statistic of 6.306 exceeds the recorded 5% upper bound of 5.176, but the output does not report rejection at the 1% level. An estimated individual tax coefficient can remain imprecise even if a joint bounds restriction is rejected. An error-correction parameter below −1 also merits care: it is compatible with adjustment that overshoots the preceding disequilibrium within the fitted dynamic equation, rather than necessarily monotonic convergence (Pesaran et al., 2001; Kripfganz & Schneider, 2020, 2023).

Any further formal break test, HAC uncertainty, or alternative WGI-vintage analysis would constitute additional work and is not presented as executed. By contrast, the matched component-equality and joint-zero tests in Table 5 have already been independently executed from the supplied. Residual diagnostics from the 24-observation ARDL cannot be treated as diagnostics for the static interaction models.

4. Results

4.1 Descriptive variation in the reconstructed archive

Table 1

Descriptive statistics for the harmonised annual archive, 1996–2024

Variable	N	M	SD	Minimum	Maximum
Real GDP growth (%)	29	4.1417	9.1791	−20.7388	28.6000

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 4 No. 3 (March) (2026)

Variable	N	M	SD	Minimum	Maximum
Tax revenue (% GDP)	29	6.7888	2.1652	3.2000	9.8985
Five-component IQ	29	−1.5312	0.1802	−1.8676	−1.2810
FDI (% GDP)	29	0.7593	1.1598	0	4.3687
Trade openness (% GDP)	29	45.7445	25.3785	0	73.3000
Population growth (%)	29	3.1582	1.1245	0.7620	6.1452

Source. Model D0; dataset descriptives for 1996–2024 (N = 29), independently compared with the supplied

Table 1 shows pronounced growth volatility: annual real GDP growth ranges from −20.7388% to 28.6000%, whereas the five-component composite governance score occupies a comparatively narrow negative interval. These values help explain why contemporaneous product-term regressions may be sensitive to extreme observations and limited institutional variation. The five-component IQ summary is background information and should not be confused with the three individual governance moderators. Likewise, zero minima in FDI and trade are not evidence of genuinely zero capital flows or external commerce until the original observation flags and data definitions are inspected. No new descriptive values have been calculated from raw data for this manuscript.

4.2 Corrected baseline ARDL results

Table 2

Dynamic baseline: long-run coefficients and error correction

Quantity	Estimate	SE / bound	p / decision
Bounds F (Case 3; k = 4)	6.306	5% upper = 5.176	Reject at 5%; not 1%
Long-run tax revenue	0.9963	1.2894	.450
Long-run FDI	1.2016	1.2036	.332
Long-run trade openness	−0.0136	0.1113	.904
Long-run population growth	1.0922	1.2702	.402
Error-correction adjustment	−1.0762	0.1956	< .001
Short-run Δtrade	0.4204	0.1592	.017

Source. Model A1; BIC-selected ARDL(1,0,0,1,0), annual growth outcome, 1997–2020 (N = 24).

The bounds statistic provides evidence against the relevant no-levels-relationship null under the recorded finite-sample 5% rule. It does not mean that each long-run explanatory variable is significant. As Table 2 shows, the estimated long-run tax coefficient is positive but imprecise (b = 0.9963, SE = 1.2894, p = .450). H1 is therefore not supported by an individual tax-coefficient test, even though the joint bounds inference differs. The negative error-correction coefficient is statistically distinguishable from zero, but its magnitude requires care in interpreting adjustment and stability. Short-run trade is positive in the particular displayed parameterisations; that is a different coefficient from the normalized long-run trade estimate. No governance term is present in Model A1.

4.3 Component-specific moderation results

Table 3

Direct side-by-side description of three separately estimated interactions

Model / component	N / period	TR × G estimate	SE	p	95% CI
C-GE · GE	29 / 1996–2024	4.6158	3.6617	.221	[−2.9781, 12.2097]

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 4 No. 3 (March) (2026)

Model / component	N / period	TR × G estimate	SE	p	95% CI
C-RL · RL	29 / 1996–2024	9.7837	10.7057	.371	[−12.4186, 31.9860]
C-CC · CC	29 / 1996–2024	−0.6010	7.8369	.940	[−16.8537, 15.6517]

Source. Models C-GE, C-RL, C-CC; separate OLS interaction regressions, 1996–2024 (N = 29).

Table 3 directly addresses H2–H4. For Government Effectiveness, the coefficient is positive but statistically imprecise (b = 4.6158, SE = 3.6617, p = .221). Its 95% confidence interval encompasses both a negative and a positive conditional association. Government Effectiveness thus remains theoretically relevant but is not empirically established as a positive moderator here. Rule of Law has a larger positive point estimate (b = 9.7837, SE = 10.7057, p = .371), yet the interval is markedly wide and crosses zero. This is not confirmation of a stronger legal-governance conditioning effect. Control of Corruption has a small negative point estimate (b = −0.6010, SE = 7.8369, p = .940), also with wide uncertainty spanning zero. Accordingly, H2, H3, and H4 receive no conventional statistical support in these regressions.

The differences between the three point estimates do not establish whether the components differ from one another. Separate standard errors quantify within-model uncertainty only. Without matched scales, jointly testable coefficients, and cross-equation covariance information, a formal statement that GE outperforms CC, or that RL is the strongest governance channel, has no inferential basis. All three p values also exceed .05 before any consideration of the three-test family. There is no statistical reason to interpret one apparent ordering of the estimates as a priority ranking of actual reforms.

The confidence intervals give more information than the simple distinction between significance and non significance. For GE, the interval [−2.9781, 12.2097] includes zero and values with opposite signs. The RL interval [−12.4186, 31.9860] is even more expansive, and the CC interval [−16.8537, 15.6517] permits both negative and positive associations around its near-zero estimate. These widths are meaningful evidence of uncertainty, not an invitation to infer that a larger positive end point indicates a desirable reform. The reported standard errors are those accompanying each interaction coefficient in the source manuscript. Without their model-specific variance estimator, temporal dependence correction, and residual diagnostics, even interval coverage should be treated as conditional on the original estimation procedure. No apparent difference in endpoints provides a direct statistical test of differences across models.

4.4 Matched component re-estimation and formal contrasts

The original separate regressions are now reproduced exactly from the uploaded analysis, including their standard errors and confidence intervals. An HC3 rerun gives product-term p values .312 (GE), .398 (RL), and .963 (CC), which do not support any separate moderation effect. Separate-model numerical ranking was never a formal test of component differences; common-year comparisons require estimating interactions in a common equation, preserving covariates, and using the covariance of their estimated differences (Brambor et al., 2006; MacKinnon & White, 1985).

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 4 No. 3 (March) (2026)

Table 4.

Replication and HC3 uncertainty in the separate component interactions

Model / component	Years; N	Covariance	b tax × G	SE	p	95% CI
C-GE	1996–2024; 29	OLS	4.616	3.662	0.221	[-2.98, 12.21]
C-GE	1996–2024; 29	HC3	4.616	4.461	0.312	[-4.64, 13.87]
C-RL	1996–2024; 29	OLS	9.784	10.706	0.371	[-12.42, 31.99]
C-RL	1996–2024; 29	HC3	9.784	11.353	0.398	[-13.76, 33.33]
C-CC	1996–2024; 29	OLS	-0.601	7.837	0.940	[-16.85, 15.65]
C-CC	1996–2024; 29	HC3	-0.601	12.716	0.963	[-26.97, 25.77]

Source. Each regression has an intercept, TR, its centered governance dimension, TR×dimension, FDI, trade openness and population growth; OLS lag order 0, residual df = 22. Estimates are recomputed from the 1996–2024 archived annual dataset and compared with the corresponding original regression output.

To test differences, a new joint model (C-J3) enters GE, RL and CC together, their three tax interaction terms, tax revenue, FDI, trade and population. Each governance measure is centered then divided by its within-archive standard deviation, so the three product coefficients refer to a one-SD governance difference. The comparison is conditional on strongly correlated component series and 18 residual degrees of freedom. It should be read as an exploratory covariance-aware exercise, not a precise test of which ministry reform should come first.

Table 5.

Joint standardized component estimates, pairwise contrasts, and omnibus restrictions

Coefficient / restriction	b / F	SE	95% CI	p OLS	p HC3
Tax × GE (per SD)	-1.273	1.919	[-5.30, 2.76]	0.516	0.577
Tax × RL (per SD)	-0.856	1.668	[-4.36, 2.65]	0.614	0.605
Tax × CC (per SD)	0.160	1.380	[-2.74, 3.06]	0.909	0.939
GE = RL	F=0.020			0.889	0.901
GE = CC	F=0.368			0.552	0.484
RL = CC	F=0.187			0.670	0.745
GE = RL = CC	F=0.232			0.795	0.764
All = 0	F=0.357			0.784	0.750

Source. All pairwise and omnibus restrictions are linear contrasts of a single OLS coefficient vector with common controls; zero-order lag. Coefficients scaled to 1 within-panel SD. Equality test F(2,18); zero-interaction test F(3,18). No multiple-testing-adjusted confirmation is claimed.

The standardized joint coefficients are negative for GE and RL and slightly positive for CC unlike some single-component point estimates. Their direction should not be interpreted substantively without acknowledging collinearity and model dependence. The equality null is not rejected under either covariance convention (p = .795 and .764). A joint zero-interactions test is also not rejected (p = .784 and .750). Diagnostics for C-J3 give Breusch–Godfrey p = .136, Breusch–Pagan p = .920 and RESET p = .657; all have limited

ISSN Online: 3007-3154
ISSN Print: 3007-3146

power at N = 29. VIF and matrix conditioning require substantive caution despite the arithmetic feasibility of the regression.

4.5 Model boundaries and evidentiary limits

The supplementary composite-IQ interaction reported elsewhere in the archive is −2.1304 (SE = 5.5997, p = .707), and an IQ quadratic test gives p = .8916. Both are non-confirmatory. They are provided here only as context for the already published composite-governance research program, not as results of a component model. A quadratic term in IQ alone is also not a test of a tax-slope threshold. No data-supported governance cutoff is established. This avoids attaching a threshold conclusion to component evidence that estimates only linear interactions.

5. Discussion

5.1 What the governance decomposition does and does not show

The three governance dimensions have distinct theoretical interpretations, but the empirical tests do not distinguish their contributions to the tax–growth association with useful precision. The study is therefore better understood as a disciplined decomposition of uncertainty than as a ranking of institutional reforms. GE concerns administrative capability, RL captures perceived legal predictability, and CC reflects perceived misuse of public power. Each could alter taxpayer compliance, investor decisions, and expenditure quality in different ways. Yet the measured interactions do not establish those mechanisms, because the intermediate outcomes tax filing behavior, dispute resolution, revenue leakage, and the productivity of public spending are not included in the regressions (Besley & Persson, 2009; North, 1990).

A particularly important mistake would be to reason that a larger positive point estimate indicates a larger growth payoff. RL has a numerically larger tax interaction than GE, but also a much larger standard error and a wide confidence interval. Its result is compatible with a broad range of conditional gradients. GE's point estimate is positive and more precise in absolute units, but not sufficiently precise to reject zero. CC's near-zero estimate has a wide interval rather than demonstrating equivalence to zero. A formal equivalence conclusion would require substantively justified equivalence bounds and an appropriate test. Thus both superiority and equivalence remain unestablished.

The observed patterns are also compatible with concurrent macroeconomic changes. Tax revenue as a GDP share and trade openness as a GDP share can move with denominator shocks. Governance perceptions may respond to conflict, public-service disruption, changes in international engagement, or observable economic performance. There are only 29 annual observations, and the most unusual years may influence fitted interaction slopes substantially. A credible causal account would need a design that isolates the timing of governance change from these common shocks, rather than relying solely on contemporaneous aggregate regressors.

These estimates also raise the possibility of measurement and model-selection uncertainty beyond the reported standard errors. Governance measures capture broad perceptions that can share information sources, which may correlate their errors and obscure distinctions among administrative, legal, and integrity functions. Meanwhile the sample size is too small to include many competing interactions, flexible lag terms, and political-period dummies without materially reducing residual degrees of freedom. This creates a trade-off: simpler separate models may be estimable but do not formally identify cross-component differences, while more comprehensive joint models may be unstable or highly collinear. Selecting the model that happens to produce the most attractive coefficient would understate this uncertainty. A publication-quality interpretation should report the chosen specifications consistently and disclose the set of alternatives considered, rather

ISSN Online: 3007-3154
ISSN Print: 3007-3146

than equating fit statistics or theoretical appeal with identification.

5.2 Measurement uncertainty and the importance of a fixed WGI vintage

WGI component scores are constructed from underlying surveys and expert assessments. The accompanying uncertainty is part of the measurement process, not a negligible technical detail. Perception-based changes are neither one-to-one measures of bureaucratic productivity nor necessarily contemporaneous responses to institutional reform (Kaufmann et al., 2011). The World Bank explicitly cautions that the 2026 WGI estimates reflect perception data and that historical values have been recalculated under revised methodology (World Bank, 2026). These characteristics make a frozen vintage, variable labels, and unit scale indispensable for reproducibility.

Updating a bibliographic citation does not establish comparability between governance-estimate vintages. Replacing a historical governance score with a value retrieved from the current WGI portal while retaining the old interaction coefficient would make the paper internally inconsistent. A defensible sensitivity exercise would retain the original series and estimates unchanged, then separately estimate each model using the 2026 WGI vintage, compare actual coefficients and confidence intervals, and disclose any model or sample changes. This exercise has not been run with the materials provided. Similarly, the estimated negative-scale governance scores must not be recast as the World Bank's absolute 0–100 scores without a new transformation and interpretation.

The composite-governance and component models also have different questions. The five-component average in the dissertation includes regulatory quality and voice and accountability, alongside GE, RL, and CC. This article asks about the three named components separately and makes no claim that a three-component average is equivalent to the five-component index. Component estimates should be located on their observed support. Arbitrary governance values particularly those never recorded in Afghanistan's historical series could yield extrapolated tax slopes without empirical relevance. Model-specific standardization and full covariance-aware slope intervals are appropriate future additions, not results already available.

5.3 Implications for research and public finance practice

Public-finance institutions remain important for reasons wider than GDP growth. Reliable assessments, transparent appeals, publication of budget execution, and controls on discretionary spending may improve fairness and accountability even where a country-level regression cannot estimate their incremental effect on real growth. The evidence presented here justifies attention to administrative records and policy-process indicators, but not a recommendation to prioritize RL over GE or vice versa on the basis of one estimated regression coefficient. Reform sequencing therefore belongs to a broader assessment of administrative constraints, distributional consequences, and implementation feasibility rather than a statistically unsupported component league table (Akitoby et al., 2020; Prichard, 2015).

A stronger future design would collect annual or higher-frequency records of tax assessments, filing completion, customs processing, audit recovery, judicial dispute resolution, procurement transparency, expenditure execution, and sectoral public-service outputs. Such measures could connect the theory of credible enforcement and efficient public inputs more directly to observed implementation. It would also be useful to obtain complete model covariance matrices, influential-observation tests, documented break dates, governance-indicator uncertainty, and pre-declared component comparison criteria. Each new specification should use a traceable code version, consistent observation set, and published data provenance. The present article does not imply that these analyses were executed.

ISSN Online: 3007-3154
ISSN Print: 3007-3146

The existing study's methodological boundaries also limit external validity. Afghanistan's institutional environment, aid dependence, and political changes make its annual series unlike an experimental policy intervention or an average cross-country reform episode. Low power and measurement error may obscure genuine associations; conversely, observed ones may reflect omitted common shocks. Statistical non-rejection does not prove an institution lacks economic importance, but it does prohibit claims of empirical confirmation. This asymmetry between theoretical plausibility and measured precision is the central lesson of the component decomposition.

Finally, a null-hypothesis test with a large p value answers a limited question. It reports insufficient evidence to reject a specific zero coefficient under a particular model; it does not measure the posterior probability that the coefficient is zero, the probability that one component dominates another, or the practical value of a policy change. Likewise, the reported 95% confidence interval is tied to assumptions about sampling, model specification, and standard errors. Because the data are a single country's historical annual series, those assumptions are especially contestable when temporal dependence, changing measurement practices, and macroeconomic shocks are incompletely represented. The most useful contribution is therefore transparency about what the observed intervals permit and what they cannot discriminate. A more demanding empirical test would identify a policy-relevant minimum change, prespecify the alternative against which it is judged, and collect sufficient data to make that distinction credible.

5.4 Independence from the related 2026 publication

The study by Kareemzada et al. (2026) already addresses tax revenue, institutional quality, and economic growth in Afghanistan using an overlapping 1996–2024 archive and ARDL–ECM framework. The overlapping observations and ARDL estimates are treated as previously reported context; this study focuses on newly computed, matched comparisons among three governance interactions.

Its distinguishable emphasis is on three named WGI components, their separate 29-observation interactions, the inability to infer cross-component differences, and governance-vintage reproducibility.

6. Conclusion

This article evaluates whether Government Effectiveness, Rule of Law, and Control of Corruption differentially condition the tax–growth relationship in Afghanistan. The three archived product-term coefficients have p values of .221, .371, and .940, respectively, and confidence intervals crossing zero. The models therefore do not establish positive moderation by any component, nor do they establish a statistically defensible ranking among components. The corrected ARDL baseline remains a separate historical model without a governance moderator; its individually imprecise tax coefficient cannot support claims of confirmed revenue-led growth.

The paper's contribution lies in making the limits of the available evidence auditable: a clear theory-to-model map, model-specific sample reporting, confidence-interval interpretation, a distinction between static moderation and dynamic adjustment, and an explicit treatment of WGI vintages and prior-publication overlap. Policy relevance follows from the need for transparent, functioning fiscal institutions not from a quantified component-specific growth dividend. Credible comparative claims will require original input files, matched model specifications, cross-equation covariance information, and independent administrative measures. Until then, the evidence supports restraint rather than ranking.

REFERENCES

ISSN Online: 3007-3154
ISSN Print: 3007-3146

Vol. 4 No. 3 (March) (2026)

Akitoby, B., Honda, J., & Primus, K. (2020). Tax revenues in fragile and conflict-affected states Why are they low and how can we raise them? (IMF Working Paper No. 2020/143). International Monetary Fund. https://doi.org/10.5089/9781513550848.001

Barro, R. J. (1990). Government spending in a simple model of endogenous growth. *Journal of Political Economy*, 98(5, Part 2), S103–S125. https://doi.org/10.1086/261726

Besley, T., & Persson, T. (2009). The origins of state capacity: Property rights, taxation, and politics. *American Economic Review*, 99(4), 1218–1244. https://doi.org/10.1257/aer.99.4.1218

Brambor, T., Clark, W. R., & Golder, M. (2006). Understanding interaction models: Improving empirical analyses. *Political Analysis*, 14(1), 63–82. https://doi.org/10.1093/pan/mpi014

Kareemzada, W., Rashid, M. U., & Shah, S. H. A. (2026). Tax revenue, institutional quality, and sustainable growth in Afghanistan (1996–2024): A dual source ARDL–ECM analysis. *Journal of Management Science Research Review*, 5(1), 374–404. https://doi.org/10.5281/zenodo.19485920

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

Kripfganz, S., & Schneider, D. C. (2020). Response surface regressions for critical value bounds and approximate p-values in equilibrium correction models. *Oxford Bulletin of Economics and Statistics*, 82(6), 1456–1481. https://doi.org/10.1111/obes.12377

Kripfganz, S., & Schneider, D. C. (2023). ardl: Estimating autoregressive distributed lag and equilibrium correction models. Stata Journal, 23(4), 983–1019. https://doi.org/10.1177/1536867X231212434

MacKinnon, J. G., & White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. Journal of Econometrics, 29(3), 305–325. https://doi.org/10.1016/0304-4076(85)90158-7

Moore, M. (2004). Revenues, state formation, and the quality of governance in developing countries. *International Political Science Review*, 25(3), 297–319. https://doi.org/10.1177/0192512104043018

North, D. C. (1990). *Institutions, institutional change and economic performance*. Cambridge University Press. https://doi.org/10.1017/CBO9780511808678

Pesaran, M. H., Shin, Y., & Smith, R. J. (2001). Bounds testing approaches to the analysis of level relationships. *Journal of Applied Econometrics*, 16(3), 289–326. https://doi.org/10.1002/jae.616

Prichard, W. (2015). *Taxation, responsiveness and accountability in sub-Saharan Africa: The dynamics of tax bargaining*. Cambridge University Press. https://doi.org/10.1017/CBO9781316275511

World Bank. (2026). Worldwide Governance Indicators: 2026 update [Data set and documentation]. https://www.worldbank.org/en/publication/worldwide-governance-indicators

Appendix A. Reproducibility and Observation Provenance

The analysis uses a harmonized annual dataset assembled from WDI and official Afghan NSIA, DAB and MoF series. These source names do not establish that all 29 observations were directly observed in both archives. For replication, the author

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 4 No. 3 (March) (2026)

should archive the source file or stable dataset vintage, download date, series code, original time unit, Afghan fiscal-year or Gregorian-calendar conversion, source priority rule, any missing-cell decision, and both pre- and post-correction values. Each adjusted observation must have an explicit reason; a zero must not be treated as equivalent to missing, and a proxy must not silently become an official observation.

Table A1

Provenance codes and mandatory documentation

Code	Interpretation	Minimum replication evidence
O	Official	Published agency series, exact date, and unaltered observation
E	Estimated	Estimation procedure, inputs, and uncertainty
P	Proxy	Proxy definition, justification, and affected years
A	Adjusted	Original and revised value, conversion rule, and rationale
Missing	Unavailable	Missingness flag; no silent replacement or artificial zero

Governance replication requires the original WGI release, negative-scale governance estimate variables rather than the revised absolute 0–100 scores, and the actual Stata commands for the three component regressions. The five-component composite is defined as IQ5 = (CC + RL + GE + VA + RQ)/5; it is not a three-component average. Archive sample-centred variables, component summary statistics, transformations, model-matched error estimators, and all e(sample) indicators. A separate 2026 WGI exercise should be named as new sensitivity analysis, with its own results tables rather than copied historical coefficients.

Appendix B. Estimation Registry and Diagnostic Record

Table B1

Unit-root screening in the source research archive

Series	ADF p	PP p
Real GDP growth	.0829	.0001
Tax revenue	.3254	.3675
Five-component IQ	.1181	.2691
FDI	.1562	.2794
Trade openness	.3272	.3938
Population growth	.0001	.0082

Table B2

Reported diagnostic tests for the corrected ARDL baseline

Diagnostic	Reported p	Scope
Breusch–Godfrey, lag 1	.9856	Serial-correlation null not rejected
Breusch–Godfrey, lag 2	.3750	Serial-correlation null not rejected
Breusch–Pagan	.7240	Homoscedasticity null not rejected
Ramsey RESET	.6534	No detected RESET misspecification
Skewness–kurtosis	.1840	Normality null not rejected
Shapiro–Wilk	.06425	Normality null not rejected

Source. Archived ARDL(1,0,0,1,0), 1997–2020, N = 24; model-specific residual diagnostics only.

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 4 No. 3 (March) (2026)

Estimation registry: D0 denotes source descriptive statistics, 1996–2024 (N = 29); A1 denotes the BIC-selected ARDL(1,0,0,1,0), 1997–2020 (N = 24); C-GE, C-RL, and C-CC denote separate described OLS interaction models, 1996–2024 (N = 29 each). The interaction equations have no distributed lag order. The descriptive and interaction results are distinguished by model and source. The archived dynamic coefficients are contextual, while the matched interaction contrasts are computed for this study. No diagnostic set for C-GE, C-RL, or C-CC was present in the provided article and none is imputed. The native Stata logs are required to confirm all coefficient-standard-error pairings, centring and sample inclusion.