Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Data Provenance and the Tax Institution Growth Nexus in Afghanistan (1996–2024): A Reproducible WDI Official Source Comparison Protocol

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 methodological article specifies a reproducible comparison of World Development Indicators (WDI) and Afghan official macro-fiscal series in analyses of taxation, institutions and growth during 1996–2024. The author-supplied harmonized archive contains 29 annual observations, while the previously estimated dynamic benchmark is a BIC-selected ARDL(1,0,0,1,0) for 1997–2020 (N = 24). The baseline reports a bounds F statistic of 6.306, a statistically imprecise long-run tax coefficient of 0.9963 ($p = .450$), and an adjustment parameter of −1.0762 ($p < .001$). These are archived model results, not outputs from alternative-source regressions. Four proposed configurations (R0–R3) would compare WDI and Afghan official GDP-growth and tax-revenue source packages using a matched annual sample, fixed covariates, and a common lag structure. Their coefficient contrasts have not been executed or verified in the supplied evidence; no data-source-driven sign reversal is claimed. The article provides a measurement crosswalk, provenance-flow diagram, year-level ledger schema, evidence classification and statistical reporting protocol. It argues that source choice should be evaluated as a transparent measurement-sensitivity decision rather than a causal instrument. The protocol establishes a transparent basis for subsequent matched-source sensitivity analysis, conditional on access to the required source records.

Keywords— Afghanistan; data provenance; World Development Indicators; measurement sensitivity; fiscal statistics; reproducibility; ARDL

1. Introduction

A regression coefficient is not independent of the observations from which it is estimated. In fragile and conflict-affected economies, the construction of those observations can be as consequential as the selection of a lag order or control variable. Afghanistan provides an important example because international macroeconomic databases coexist with national accounts, central-bank statistics and fiscal administrative reports whose coverage, reporting calendars and historical revisions need not coincide. Apparent annual continuity can conceal estimates, proxy

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

observations or adjustments applied retrospectively. These characteristics complicate any inference that treats a harmonized macroeconomic spreadsheet as a collection of uniformly observed economic facts. Jerven (2013) describes the wider developmental significance of statistical revisions, while Christensen and Miguel (2018) identify documented inputs and reproducible analysis as central to credible economic research. The central question is not whether World Development Indicators (WDI) or Afghan official accounts are intrinsically superior. It is how an analyst can determine whether a fiscal–growth result depends on source selection without conflating a change in measurement with a change in sample, estimation method or economic construct. A tax series labeled 'percent of GDP' may use a revenue numerator compiled under one government-finance perimeter and an output denominator benchmarked under another national-accounts release. A growth series may likewise change when output levels are revised. Replacing one reported series is therefore not necessarily a neutral substitution of equivalent measurements. The appropriate response is an explicit crosswalk that documents source identity, coverage, scale, revision vintage and the rule that maps each input into an analysis-ready observation.

This paper develops and specifies such a protocol for Afghanistan over 1996–2024. The corrected dissertation-derived research archive contains 29 annual records and a separately reported BIC-selected ARDL(1,0,0,1,0) estimated over 1997–2020 (N = 24). It also provides a descriptive near-zero GDP-growth–tax correlation for its harmonized master series. What the supplied archive does not document is a matched four-run comparison that alternates WDI and official GDP and tax inputs while holding estimation years, equations and lag order fixed. The paper does not fill that gap with illustrative or invented coefficients. Instead, it treats the existing baseline as contextual archived evidence and presents the WDI–official comparison as a reproducible, pre-specified research protocol.

Three questions structure the contribution. RQ1 asks which provenance and classification decisions must be visible for annual WDI and official series to be meaningfully compared. RQ2 asks how a fixed-design source-substitution experiment should distinguish substitution of the real-output measure, substitution of the fiscal measure, and their joint substitution. RQ3 asks what the existing Afghanistan archive actually establishes before the four planned comparisons are run. The corresponding objectives are to define a machine-readable year-by-variable ledger, specify a paired-sample factorial comparison with explicit eligibility rules, and separate verified published information, archived statistical claims, arithmetic computations and unexecuted research tasks. These are design and audit objectives; they do not assert new causal tax effects.

The proposed contribution differs from a conventional tax–growth estimation paper. A published coauthored study already discusses Afghanistan's fiscal, institutional and growth relationships using an ARDL–ECM framework (Kareemzada et al., 2026). The present work does not relabel that paper's findings or the corrected thesis coefficients as new country-level discoveries. Its distinct scholarly claim is methodological: an auditable source-comparison architecture can expose when a reported robustness exercise changes more than one empirical object at once. The value of this claim depends on its transparent status as a protocol, on the usability of its data dictionary,

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

and on eventual implementation—not on a predetermined sign reversal.

2. Literature Review and Analytical Framework

2.1. Fiscal capacity, data generation and measurement validity

Fiscal-capacity theory explains why tax receipts should be interpreted in their institutional setting. Besley and Persson (2009) model the complementarity of investments in fiscal and legal capacity, emphasizing that observed policy outcomes are constrained by previously developed institutions. Consequently, an increase in a tax share can reflect changes in the taxable base, compliance, fiscal administration, collection coverage or the GDP denominator. The theoretical relevance of taxation to development does not establish which of these components generated a given national-account number. For empirical research, the first requirement is to identify what a fiscal indicator measures and the agencies, accounting frameworks and revisions that produced it.

WDI standardizes indicator definitions and supports cross-country comparison, but standardization is not equivalent to direct collection of every underlying observation. For instance, the WDI tax-revenue share is based on government-finance concepts and can draw on IMF government-finance reporting. The WDI GDP-growth indicator reflects constant-price national accounts and may incorporate official-country or World Bank estimates. The World Bank's indicator metadata identify the substantive measure and broad source category but are not a substitute for the Afghanistan-specific dated extract underlying an archived coefficient (World Bank, n.d.-a, n.d.-b). An Afghan fiscal bulletin may offer better institutional detail while employing a different fiscal year or revenue perimeter. Both documents can be legitimate for their stated purposes without being interchangeable for a fixed statistical estimand.

A critical implication follows from using ratios. Write tax revenue as T and nominal GDP as Y. The reported tax measure is 100T/Y. If Y is rebased or revised, the ratio changes even if cash collections T do not. The same denominator may enter trade openness, foreign direct investment ratios and other covariates; their joint movement can reflect shared measurement construction. GDP growth is a transformation of a real-output series, not simply the denominator of the fiscal ratio. Hence an experiment that changes the output series but not the denominator used in the tax ratio studies a different contrast from an experiment that reconstructs both using the same official GDP vintage. Both are potentially informative, provided the researcher labels which is being conducted. Neither amounts to using source choice as an instrumental variable.

2.2. Statistical capacity, provenance and reproducibility

Jerven's (2013) work on African development statistics demonstrates how changes in statistical capacity and GDP benchmarks can alter narratives about economic levels and trends. The precise pathways in Afghanistan require separate evidence, but the methodological lesson transfers: metadata and the data-producing process belong inside the empirical argument. Christensen and Miguel (2018) argue for transparent research workflows that permit replication, independent scrutiny and assessment of specification choice. Gelman et al. (2020) similarly emphasize understanding how measurements are constructed before turning them into regression covariates. For a short annual time series, a single reconstructed observation may affect the sample mean, estimated persistence, information-criterion choice and coefficient uncertainty.

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

A provenance register can make these decisions auditable if it is more detailed than a list of source institutions. The minimum unit of documentation is a variable–year–source record, with fields for download time, publication release, data vintage, calendar or fiscal reference period, original observation, unit, transformation formula, administrative adjustment, coverage quality and an explicit missingness code. A later chosen 'research value' is derived from those source records and should never overwrite them. A distinction among an observed zero, an unavailable figure, an estimate and a proxy is necessary because replacing unobserved early trade or FDI with numeric zeros can affect both regression fit and substantive interpretation.

2.3. Governance indices and retrospective revisions

The thesis's institutional-quality series is an author-constructed equal-weighted mean of five Worldwide Governance Indicators (WGI): Government Effectiveness, Rule of Law, Control of Corruption, Voice and Accountability, and Regulatory Quality. It excludes Political Stability and Absence of Violence/Terrorism. This composition has analytical significance and must not be described as one of the World Bank's six official WGI dimensions. Earlier work documents uncertainty inherent in the perception-based estimates (Kaufmann et al., 2011). The 2026 WGI release recalculated historical observations back to 1996 and distinguishes standardized governance estimates from a separate anchored 0–100 score (World Bank, 2026). A regression reproducing older thesis coefficients must preserve the actual original governance vintage and score convention.

The fiscal-source comparison should initially hold this constructed IQ vector fixed across configurations. Otherwise an apparent change in a tax coefficient may reflect simultaneous alterations in fiscal data and governance measurement. A subsequent governance-vintage analysis can be justified, but it is a separate sensitivity family, with its own data manifest, model IDs and output. Because no matched WGI-vintage rerun accompanies this manuscript, no empirical effect of the 2026 revision is claimed.

2.4. Small-sample dynamic inference and what source comparisons test

The ARDL bounds framework tests for a conditional levels relationship under specified integration assumptions, lags and deterministic terms (Pesaran et al., 2001). Kripfganz and Schneider (2023) show how the Stata ARDL implementation expresses the same selected equation in levels and error-correction forms and supplies finite-sample-oriented bounds inference. These tools do not render a coefficient immune to measurement revisions or guarantee causal identification. In an effective sample of only 24 annual observations, source replacement may change residual serial correlation, fitted lag dependence and stability. A scientifically interpretable comparison must first reproduce the selected benchmark and then change a controlled feature, while reporting exactly how that feature was changed.

A comparison of two coefficients from the same matched years is also statistically dependent. Subtracting their point estimates describes observed estimator sensitivity, but the standard error of the difference requires cross-model covariance or a defensible paired resampling method. Confidence intervals for separate coefficients do not themselves constitute a confidence interval for their contrast. The present protocol specifies this inference requirement but does not purport to calculate it. Strong source sensitivity, weak sensitivity and near invariance are all admissible outcomes of a

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

properly executed protocol. Defining a dramatic sign reversal as the expected finding would invite specification searching rather than a falsifiable design.

2.5. An observation-to-estimate framework

Figure 1 separates the underlying data generation, the deterministic harmonization process and the econometric layer. Solid paths represent procedural operations that can be reproduced once the requisite source files are frozen. The final dashed decision stage represents a comparison that is specified but not yet executed. This visual distinction prevents a process diagram from being mistaken for evidence that the data sources caused a particular coefficient change. The framework has broader applicability because it supplies a testable chain from original values to a reported table cell: a referee can identify which link is documented and which remains absent.

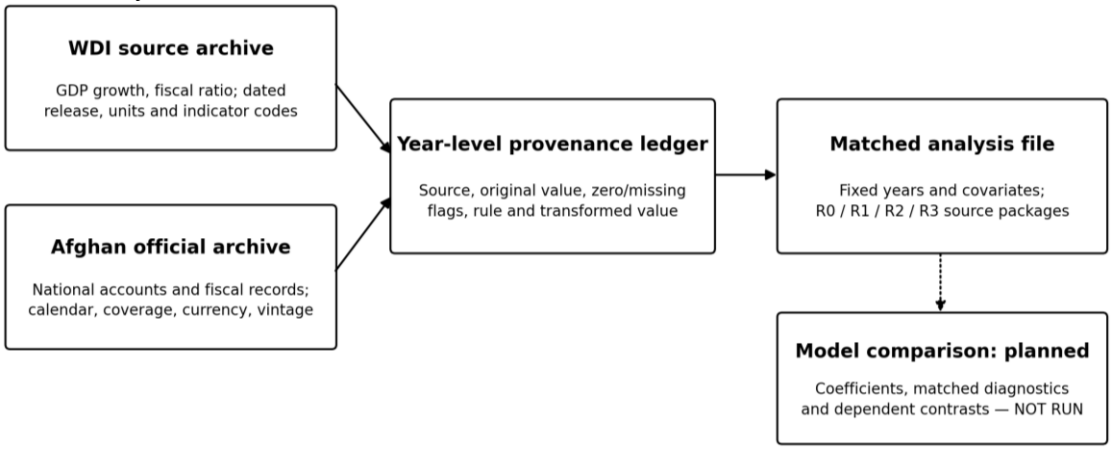

Figure 1. Evidence hierarchy and WDI–official source comparison workflow

source. The figure presents a conceptual workflow; no matched-source coefficient comparisons have been estimated.

3. Data, Provenance and Source-Comparison Protocol

3.1. Research archive, measurements and sample boundaries

The archived annual master dataset covers 1996–2024, a nominal 29 calendar-year record. The growth outcome is annual real GDP growth in percentage points; tax revenue, FDI and trade openness are recorded as percentages of GDP, and population growth is an annual rate. IQ is the previously described constructed five-component governance average. These are concepts and units reported in the supplied study, not proof that every raw year was directly measured under an invariant statistical standard. Official Afghan materials cited in the thesis include Da Afghanistan Bank, the National Statistics and Information Authority, and the Ministry of Finance. The exact file releases and observation-level precedence rules are have not yet been independently documented in the available replication materials and therefore remain replication metadata to be supplied.

The archive should be divided into four layers: unmodified WDI downloads and original Afghan published files; an extraction-and-reconciliation ledger; a harmonized analysis dataset; and executable model commands with complete postestimation logs. Each layer should have a recorded version date and checksum. The harmonized dataset must retain both source alternatives, the final selected value and a reason for the

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

selection. Temporal inclusion flags are model-specific. The primary BIC-selected ARDL has an effective 1997–2020 sample with N = 24, whereas the master descriptive records span 29 years. It is therefore incorrect to describe the primary dynamic coefficient as an estimate using all 29 years or the post-2021 period.

3.2. Measurement crosswalk and denominator integrity

Before regression, the tax series must be mapped to a defined government-finance numerator and a nominal-GDP denominator. The source-specific growth outcome must similarly be reconstructed or validated against the stated real-output measure and timing basis. A pair of similarly named indicators is not equivalent if one covers budgetary central government while another includes broader public entities, if a fiscal-year revenue numerator is assigned to a calendar-year GDP denominator without a bridge, or if one source uses a revised GDP base. The crosswalk therefore treats country, source agency, series code, governance/fiscal coverage, current-versus-constant price, national-currency unit, fiscal calendar, revisions and zeros as substantive metadata. Table 1 specifies the fields needed to establish comparability.

Table 1

Minimum measurement crosswalk for two-source fiscal comparisons

Concept	International-source field	Official-source field	Primary reconciliation gate
Real GDP growth	WDI annual constant-price growth	DAB / NSIA growth series	Price base; calendar year; release
Tax ratio	WDI tax / GDP ratio	MoF taxes and selected nominal GDP	Government perimeter; grants; ratio denominator
Controls	WDI FDI, trade, population	If replaced, define separately	Freeze original covariate values for R0–R3
IQ index	Historical five-component WGI mean	No official equivalent assumed	Freeze WGI estimate vintage and index formula
Missingness	Null / estimated / imputed if identified	Zero / unavailable / proxy	Do not equate an absent value with zero

source. Illustrative WDI indicators for GDP growth and tax revenue as a share of GDP are NY.GDP.MKTP.KD.ZG and GC.TAX.TOTL.GD.ZS. The underlying study-specific releases and extract dates have not been independently authenticated (World Bank, n.d.-a, n.d.-b).

The dataset must also distinguish source value from transformation. Suppose an official revenue figure is reported in AFN billions for a fiscal period and a candidate WDI ratio is already expressed in percent of calendar-year GDP. Directly replacing the ratio with the currency figure is a unit error. A valid comparison needs a chosen compatible numerator, denominator and time mapping, with the resulting formula retained in the ledger. Where no defensible mapping is possible, the year is unavailable for the primary matched-source experiment. Such omission may reduce sample size, but suppressing it through undocumented interpolation would transform a source comparison into an

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

unacknowledged imputation exercise.

3.3. Governance definition and source-vintage freeze

Define IQ_t = (CC_t + RL_t + GE_t + VA_t + RQ_t)/5, with all components extracted from the same archived WGI release and using the same underlying estimate scale. The index is not a direct measurement of tax compliance, fiscal-productivity improvement or customs performance. For the R0–R3 test, the IQ series is fixed; the model versions differ only in the predesignated GDP-growth and tax-revenue source packages. A different historical WGI release must receive a separate sensitivity identifier. If later governance interactions are evaluated, the centring reference and observed score support also must be frozen across source configurations to prevent a transformation change from masquerading as a data-source finding.

3.4. Paired R0–R3 factorial design

Let G^W_t and G^O_t denote harmonized WDI and official real-GDP-growth measures, respectively; let TR^W_t and TR^O_t denote tax ratios constructed under their declared fiscal-source and GDP-denominator conventions. The four planned configurations are R0 = (G^W, TR^W), R1 = (G^O, TR^W), R2 = (G^W, TR^O), and R3 = (G^O, TR^O). A superscript indexes a source package, not the truth of the corresponding value. The primary pairwise experiment uses the intersection sample S* where all four required input series and all unchanged controls are observed and definitionally comparable. R1–R3 have not been documented as executed in the supplied materials. Their role is to make source choices explicit and falsifiable, not to populate a table of nonexistent estimates.

R0: (Gw, TRw); R1: (Go, TRw); R2: (Gw, TRo); R3: (Go, TRo). (1)

Table 2 is a design registry rather than an empirical results table. The proposed comparison fixes the observation years, included covariates, transformations, deterministic terms, estimator, lag structure, standard-error convention, and pre-analysis decision rule. A flexible design that reselects lags after each source substitution may be used later to examine model-selection dependence, but it answers a different question. The fixed-lag comparison isolates source substitution conditional on one model. It should be reported first. If too few paired observations survive to support the chosen dynamics, the proper conclusion is that the planned test is infeasible under the stated restrictions, not that a favorable coefficient can be recovered by silently relaxing them.

Table 2

Pre-specified R0–R3 source-substitution protocol

Run ID	Growth outcome	Tax measure	Common setting	Status
R0	WDI	WDI	S*; fixed Z, lags and VCE	Unexecuted
R1	Official	WDI	Same S*, Z, lags and VCE	Unexecuted
R2	WDI	Official	Same S*, Z, lags and VCE	Unexecuted

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

Run ID	Growth outcome	Tax measure	Common setting	Status
R3	Official	Official	Same S*, Z, lags and VCE	Unexecuted

source. "WDI" and "official" denote completely documented source packages, including necessary numerator and denominator definitions. S is the all-source paired-year intersection. This is a design table, not a table of fitted coefficients.*

3.5. Conditional models and source-sensitivity estimands

For a general fixed-design static model, the tax–growth association can be expressed as G_t^(s) = α_s + β_s TR_t^(s) + γ′Z_t + ε_t^(s), where Z includes unchanged controls such as FDI, openness, population growth and, if required by the chosen specification, a frozen IQ index. For a dynamic sensitivity design, the benchmark lag structure must first be fixed using prespecified criteria and then applied identically to the source configurations. A tax coefficient is therefore meaningful only alongside its conditional model and sample. The primary descriptive source effect is the paired change in a named estimand θ_s, written Δ_s = θ_s − θ_R0. Such a difference is a change in an estimated association, not an economic causal response to statistical-source selection.

$$G_t^{(s)} = \alpha_s + \beta_s\, TR_t^{(s)} + \gamma' Z_t + \varepsilon_t^{(s)}. \quad (2)$$

$$\Delta_s = \theta_s - \theta_{R0}, \quad s \in \{R1, R2, R3\}. \quad (3)$$

The four-run design also supports a descriptive interaction contrast in measurement sensitivity, D = (θ_R3 − θ_R2) − (θ_R1 − θ_R0). This is a difference of coefficient changes under alternative source packages; it is not a fiscal or institutional interaction in the economy. Estimation should account for correlation among models fitted on the same years and the limitations of time-series resampling in very short samples. Percent changes should not be emphasized when the reference coefficient is near zero. Inference ought to emphasize estimated differences, their intervals where justified, influence diagnostics and a full record of omitted years, rather than significance stars alone.

$$D = (\theta_{R3} - \theta_{R2}) - (\theta_{R1} - \theta_{R0}). \quad (4)$$

3.6. Archived ARDL baseline: role and interpretation

The controlling archived regression is a BIC-selected ARDL(1,0,0,1,0) for 1997–2020 (N = 24), with contemporaneous tax revenue, FDI and population growth, current and lagged trade openness and lagged GDP growth. Its reported levels form is g_t = a + ρg_(t−1) + bTR_t + cFDI_t + d0OPEN_t + d1OPEN_(t−1) + ePOP_t + v_t. Institutional quality and political-period indicators are not present. The reported conditional tax long-run coefficient is b/(1 − ρ), while the openness long-run numerator uses d0 + d1. The equivalent error-correction representation depends on that exact lag structure and software convention. The baseline model is a separate archived result; it is not presumed to equal R0 unless a verified source-identity comparison establishes that equivalence.

$$g_t = a + \rho g_{t-1} + bTR_t + cFDI_t + d_0 OPEN_t + d_1 OPEN_{t-1} + ePOP_t + v_t. \quad (5)$$

Bounds inference for that model is conditional on its integration admissibility, deterministic case and postestimation critical values. It tests a joint levels restriction, not the causal effect of taxation or the significance of a single normalized slope. Further, a growth-rate dependent variable means normalized coefficients describe conditional

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

percentage-point associations, not elasticities of the GDP level. The source manuscript reports an error-correction parameter below −1; this may correspond to fitted overshooting under the maintained dynamics but must not be called recovery of lost national output. No newer-source coefficient, residual diagnostic or bounds decision can inherit the archived ARDL result without estimation.

3.7. Provenance status and evidence classification

The manuscript classifies evidence into four categories. A denotes numbers attributed to the corrected dissertation or Stata summaries but not independently regenerated from the original code in this task. B denotes documentary facts and bibliographic metadata checked against institutional or publisher pages. C denotes arithmetic calculations obtained deterministically from stated inputs. D denotes unexecuted research operations. An R0–R3 design row is D even when its intended inputs come from reputable agencies. A reported baseline regression remains A until the author supplies code and raw data allowing its independent reproduction. The origin class should appear in table notes and file manifests so that reading an attractive table does not imply stronger verification than was performed.

Table 3

Numerical-evidence classes and permissible language

Class	Origin	Permissible reporting	Release condition
A	Thesis/Stata summary	"Archived output reports…"	Replicate raw panel and exact command
B	External publisher/agency record	Verified bibliographic or indicator fact	Archive URL, title and retrieval date
C	Arithmetic from disclosed numbers	Transparent constructed quantity	Publish inputs and formula
D	Proposed analysis	Protocol only, no p values	Run, validate, archive model outputs

source. The class indicates provenance rather than statistical reliability or causal credibility.

3.8. Replication acceptance rules

A regression may enter a completed sensitivity-results table only if its model ID resolves to a frozen input file, a checksum, the exact eligible years, transformation script, command, version of Stata and user-written routines, coefficient table, variance estimator and postestimation diagnostics. The four runs must share S* and fixed model terms. The researcher should then report leave-one-year-out influence, checks excluding disputed early zero-coded entries, accounting-denominator variants and a separately identified alternative-vintage WGI exercise. Every change must receive its own model suffix. Negative results and computational failures must remain visible; 'not estimated,' 'failed specification,' 'data unavailable' and 'not statistically significant' are distinct scientific states.

Figure 1 describes a proposed workflow; the reviewed materials do not establish that

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

all input files and executable components required for R0–R3 have been deposited. The materials needed for independent scrutiny would include source snapshots where redistribution is permitted, source citations where licensing restricts deposition, one machine-readable input ledger, the frozen analysis-ready panel, model-specific flags, executable scripts and model-matched diagnostic outputs.

4. Archived Evidence and the Boundary of Empirical Claims

4.1. The harmonized dataset describes one constructed series

The author-supplied summary reports 29 observations for 1996–2024. Real GDP growth has a mean of 4.1417% and standard deviation of 9.1791 percentage points; tax revenue averages 6.7888% of GDP (SD = 2.1652), and the five-component IQ index averages −1.5312 (SD = 0.1802). FDI averages 0.7593% of GDP, openness 45.7445% of GDP and population growth 3.1582%. These describe one harmonized master dataset. The recorded minimum of zero for FDI and openness requires historical source confirmation before it is taken to mean observed economic inactivity. The summary does not establish the content of the four prospective paired-source panels, which require independent raw-source extraction and matched-year eligibility checks.

Table 4

Archived harmonized sample descriptives, 1996–2024 (N = 29)

Variable	Mean	SD	Minimum	Maximum
Real GDP growth (%)	4.1417	9.1791	−20.7388	28.6000
Tax revenue (% GDP)	6.7888	2.1652	3.2000	9.8985
Five-part IQ index	−1.5312	0.1802	−1.8676	−1.2810
FDI (% GDP)	0.7593	1.1598	0.0000	4.3687
Trade openness (% GDP)	45.7445	25.3785	0.0000	73.3000
Population growth (%)	3.1582	1.1245	0.7620	6.1452

source. Early FDI and trade zero entries require confirmation as measured zeros rather than gap codes.

The master correlation between GDP growth and tax revenue is reported as r = −0.0002, with p = .9993, whereas tax revenue and trade openness correlate at r = .8717. The first observation does not establish the absence of a conditional fiscal relationship; the second cannot show that taxes are collected through customs or that customs capacity has improved. Both are particular descriptive relationships in an archived harmonized panel. The user-supplied source manuscript also mentions different positive tax–growth correlations in older draft tables. Without the underlying year-by-year datasets, these differences cannot be attributed to a WDI–official swap, a reporting correction or an independent replication. Their unresolved origin is a substantive audit finding.

4.2. The dynamic benchmark is not a source-swap experiment

Table 5 records selected M01 estimates as archived claims. The finite-sample bounds F statistic is 6.306, above the reported 5% I(1) upper critical value of 5.176. The normalized long-run tax estimate is 0.9963 (SE = 1.2894, p = .450; 95% CI [−1.7241, 3.7166]). The adjustment estimate is −1.0762 (SE = 0.1956, p < .001), and the differenced openness coefficient is 0.4204 (SE = 0.1592, p = .017). These statistics

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

belong to the 1997–2020 ARDL and cannot be transferred to a 1996–2024 model, an IQ interaction or a source substitution. Rejection of the bounds null does not override the tax coefficient's imprecision or certify parameter stability.

Table 5
Selected archived ARDL benchmark M01, 1997–2020 (N = 24)

M01 term / test	Coefficient or test	SE	p	95% CI / qualification
Bounds F (Case 3)	6.306	—	5% rejection	5% upper I(1) = 5.176
Tax, long-run	0.9963	1.2894	.450	[−1.7241, 3.7166]
FDI, long-run	1.2016	1.2036	.332	[−1.3378, 3.7410]
Openness, long-run	−0.0136	0.1113	.904	[−0.2483, 0.2212]
Population, long-run	1.0922	1.2702	.402	[−1.5876, 3.7720]
Error correction	−1.0762	0.1956	< .001	[−1.4889, −0.6636]
Δ openness	0.4204	0.1592	.017	[0.0846, 0.7561]

source. Class A: separate archived BIC-selected ARDL(1,0,0,1,0), GDP-growth outcome, k = 4. These are not R0–R3 source-swap results.

The same family of documents reports mixed unit-root-screening evidence and concerns about stability. The growth variable's ADF and Phillips–Perron level decisions differ, and the CUSUM path is reported to cross a nominal confidence boundary while CUSUMSQ does not. Residual checks with non rejecting p values cannot resolve those distinct issues. Exact diagnostic commands must be matched to the selected M01 model rather than older ECM specifications. No CUSUM plot has been reconstructed or attached to this article as newly verified evidence. For a provenance-method paper, the methodological conclusion is that a future R0–R3 run must reproduce all relevant diagnostics for each alternative measurement configuration, not reuse M01's tests.

4.3. Status of the four proposed source configurations

The supplied manuscript contains no completed matched R0–R3 coefficient, standard-error, p-value or confidence-interval matrix. Consequently, there is no observed result that swapping official for WDI GDP growth reverses the tax coefficient, makes moderation significant or alters bounds rejection. A statement that published source alternatives exist does not demonstrate that they are defined compatibly, overlap for sufficient years or yield a converged common specification. Table 2 intentionally displays 'not executed' for each proposed model; its absence of coefficient fields is scientifically preferable to a fabricated quantitative finding. The paper's empirical contribution at this stage is a bounded baseline audit coupled with an executable research design awaiting source-complete implementation.

This evidentiary distinction applies to governance moderation and political-transition claims as well. The selected ARDL excludes IQ; a separate full-sample static

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

interaction is reported elsewhere in the thesis, but its coefficient does not become part of the ARDL when the two are discussed together. No new fiscal-by-governance interaction is estimated for R0–R3 here. Similarly, political categories or administrative transitions may explain why source coverage differs, but without a dated coding ledger and appropriate design they do not identify a causal effect of political authority on measurement or economic activity.

5. Discussion

5.1. A contribution grounded in comparability rather than dramatic estimates

The strongest defensible claim is methodological. This study transforms a vague instruction to 'check alternative sources' into a set of decisions that can be independently inspected: establish source coverage, reconcile constructs, freeze a paired sample, fix a model, replace one source package at a time, compute dependent coefficient contrasts, and preserve every changed or missing observation. The sequence is valuable in a fragile setting because measurement discontinuities can coincide with real disruptions, and ignoring either side risks attributing statistical changes to inappropriate causes. It does not guarantee a nonzero sensitivity effect. An eventual null result would still be informative if the paired source products and estimation outputs were genuinely replicated.

This approach extends the lesson of Jerven (2013) by connecting historical statistical revisions to model-level traceability. It also operationalizes the transparency agenda articulated by Christensen and Miguel (2018): empirical claims become testable when readers can connect a final coefficient to code and inputs. The contribution is strongest when accompanied by an openly specified ledger and software-independent arithmetic crosschecks, rather than a collection of narrative descriptions. Because the present manuscript lacks the executed source swaps, the status of the contribution is protocol specification and documentary audit, not an experimental demonstration of numerical sensitivity.

5.2. Why denominator integrity is an especially consequential gate

Tax revenue relative to GDP combines administrative accounts and macroeconomic output measurement. Revisions to the GDP denominator can induce apparent changes in fiscal mobilization without any change in cash collected. In a multivariate regression, the same revision may affect the growth outcome and ratios for tax, investment and trade. A simplistic 'official-GDP swap' that changes all related fields simultaneously cannot isolate whether coefficient movement arose from the output outcome, tax ratio or other controls. The two-factor R0–R3 layout helps organize this problem but does not solve it unless construction rules are explicit. A source crosswalk should retain separate nominal revenue numerators and output denominators and state whether the paired experiment changes the output only, the ratio only or a consistent set of related measures.

In addition, fixing years creates a legitimate external-validity cost. If official observations are missing precisely during periods of conflict or institutional interruption, the matched sample S* may exclude those years. The four runs can then be compared conditionally on observed overlap, but their results cannot automatically

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

be generalized to the entire 1996–2024 span. A second, explicitly separate analysis may evaluate sample-coverage sensitivity. It should not be folded into the source-only contrast. The manuscript's proposed year-eligibility field makes this distinction visible and allows readers to see which periods remain unidentified by the data rather than concealed by interpolation.

5.3. What the historical archive can and cannot validate

The corrected historical ARDL offers a useful reference for model provenance: it identifies the reported dependent variable, sample window, selected lag order and coefficients. That record demonstrates the difference between observing a levels-test rejection and estimating a precise individual tax slope. It does not validate a 29-year tax effect, a post-2021 forecast, fiscal mediation through customs, or a matched WDI–official sensitivity. Nor does the separate 2026 coauthored publication authorize a later manuscript to report its estimates as if they were newly identified. The differences between the published paper's figures and the corrected thesis results should be acknowledged as model-provenance questions until original commands and samples can be reconciled (Kareemzada et al., 2026).

The governance series supplies a further warning. Updating a perception-based composite is not a cosmetic operation when older years are recalculated. If a fiscal-source swap and a WGI-vintage change occur together, no simple interpretation can allocate a coefficient difference between them. The correct design holds IQ fixed at the historical thesis vintage for the initial swap tests and later evaluates a WGI revision in its own labelled block (World Bank, 2026). This also avoids attributing fiscal administrative productivity to a broad governance perception index whose underlying sources may themselves respond to contemporary events.

5.4. Scope, originality and transferability

A protocol article can be a legitimate journal contribution when a journal publishes reproducibility designs, research methods or registered reports, but it must be labelled accurately. The four R0–R3 source-comparison estimates have not been produced; the present contribution is the protocol for conducting and documenting those comparisons. The contribution must be assessed against the 2026 coauthored tax–growth publication, the author's related governance, ARDL, structural-break and customs manuscripts, and any simultaneous submissions.

The approach may be transferable to other fragile or low-information country settings that have both internationally harmonized and national administrative series. What transfers is the method, not an Afghan coefficient: preserve dated sources, separate numerator from denominator, document missingness, estimate only on matched years and treat coefficient contrasts as dependent. Independent use could reveal substantial source sensitivity or near invariance. Neither outcome is prejudged here. Broader methodological value would be enhanced by a deposited, usable template and a completed application once the official-source panel and original executable files become available.

5.5. Uncertainty, negative findings and pre-specified comparisons

A rigorous source experiment requires more than four coefficient columns. It should disclose changes in effective sample coverage and in the variance of the underlying series, but those are not substitutes for coefficient uncertainty. Each fitted regression

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

should report the response, exact regressor list, year range, effective N, lag structure, standard-error estimator, confidence interval and model-matched residual diagnostics. The comparison should also record the file and command from which those numbers were exported. Inference about source-induced coefficient changes should report the coefficient contrast and its uncertainty interval; comparing significance thresholds across separate regressions is insufficient. In short annual panels with cross-time dependence, ordinary independent-sample t comparisons of two estimates calculated from the same years are not valid by default.

A disciplined analysis could preregister a small number of summary contrasts before accessing the official replacements. One primary contrast might be the tax long-run coefficient under matched source versions; another could be the difference between forecast errors from otherwise identical specifications. The tolerance for describing a coefficient as substantively stable should be tied to a stated economic scale and uncertainty, not selected after examining the direction of the results. Nonconvergence, a deficient lag specification, an inadmissible integration order or loss of too many source-matched years would be reported as a failed or ineligible run. Those outcomes are scientifically informative because they identify where the proposed data comparison cannot support the intended inferential question. They must not be recoded as evidence of no source sensitivity.

5.6. Requirements for empirical implementation

The requirements for upgrading this protocol to an empirical results manuscript are concrete. First, the author must obtain dated WDI and Afghan official series for every source package and publish their conceptual crosswalk. Second, all four configurations must be constructed using a frozen intersection sample without changing other covariates or lags. Third, the estimated models must be accompanied by code, input hashes, complete model tables and diagnostics. Fourth, the source contrasts must be tested with appropriate accounting for shared observations, and sensitivity to disputed zeros, calendar allocation and leverage years must be documented. Fifth, the registered interpretation should be followed even if the results show stable coefficients or lack sufficient precision to distinguish sources.

The proposed design distinguishes methodological transparency from empirical verification. Without dated paired annual data, frozen transformation rules, model-matched results, and replicable uncertainty estimates, no conclusion can be drawn about source-induced changes in the tax coefficient. The protocol may be transferred to other settings with competing statistical systems, provided the source-definition crosswalk is adapted to the relevant institutions.

6. Conclusion

Reliable fiscal–growth inference requires a transparent account of how observations were constructed. This paper specifies a WDI–official comparison protocol for Afghanistan's 1996–2024 research archive and separates a previously reported 1997–2020 ARDL from the unexecuted four-run source-sensitivity experiment. It provides a reproducibility ledger schema, matched-sample eligibility rules, source-factorial estimands, a model-provenance hierarchy and explicit conditions for publishing numerical coefficient contrasts. The archived growth–tax correlation and dynamic estimates are reported as existing results, not as evidence that official-source

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

substitution has already altered tax coefficients. The current finding is thus an independently inspectable research design, not a confirmed causal or source-reversal claim. A full empirical demonstration requires dated paired annual source files, frozen transformations, matched R0–R3 estimation outputs, time-dependence-aware contrast uncertainty, diagnostics and documented prior-publication independence. The four R0–R3 comparisons remain unestimated; empirical evaluation requires matched annual data and reproducible estimation outputs.

References

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

Christensen, G., & Miguel, E. (2018). Transparency, reproducibility, and the credibility of economics research. *Journal of Economic Literature*, 56(3), 920–980. https://doi.org/10.1257/jel.20171350

Gelman, A., Hill, J., & Vehtari, A. (2020). *Regression and other stories*. Cambridge University Press. https://www.cambridge.org/highereducation/books/regression-and-other-stories/DD20DD6C9057118581076E54E40C372C

Jerven, M. (2013). *Poor numbers: How we are misled by African development statistics and what to do about it*. Cornell University Press.

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. (2023). ardl: Estimating autoregressive distributed lag and equilibrium correction models. *The Stata Journal*, 23(4), 983–1019. https://doi.org/10.1177/1536867X231212434

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

World Bank. (n.d.-a). GDP growth (annual %) [Indicator NY.GDP.MKTP.KD.ZG; World Development Indicators metadata]. https://databank.worldbank.org/metadataglossary/world-development-indicators/series/NY.GDP.MKTP.KD.ZG

World Bank. (n.d.-b). Tax revenue (% of GDP) [Indicator GC.TAX.TOTL.GD.ZS; World Development Indicators metadata]. https://databank.worldbank.org/metadataglossary/world-development-indicators/series/GC.TAX.TOTL.GD.ZS

World Bank. (2026). Worldwide Governance Indicators: 2026 update [Data set and

Online ISSN 3007-3197
Print ISSN 3007-3189

documentation]. https://www.worldbank.org/en/publication/worldwide-governance-indicators