

Do Governance Thresholds Survive Measurement Uncertainty? A Bayesian Errors-in-Variables Dynamic Panel Analysis

Nitiphong Songsrirote

`nitiphong.s@msu.ac.th`

Maharakham University

Research Article

Keywords: governance, measurement error, threshold model, Bayesian dynamic panel, Worldwide Governance Indicators, economic growth

Posted Date: October 21st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10924430/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Do Governance Thresholds Survive Measurement Uncertainty? A Bayesian Errors-in-Variables Dynamic Panel Analysis

Abstract

Reported governance thresholds are policy-relevant only if the institutional score used to assign regimes is measured reliably. However, threshold applications commonly condition on Worldwide Governance Indicators (WGI) point estimates, even though the producer reports country-year-specific uncertainty. This study evaluates whether governance nonlinearity survives when measurement uncertainty, latent institutional dynamics, and growth dynamics are estimated jointly. Using 3,485 country-year observations for 167 economies during 2002-2023, the analysis estimates a six-indicator Bayesian errors-in-variables dynamic panel with WGI standard errors and a smooth regime transition in annual real gross domestic product (GDP) per capita growth. The locked final audited latent smooth-transition model (M4R) yielded a threshold median of 0.203 on the rule-of-law-anchored latent governance scale (95% credible interval: -1.295 to 1.152). The lower- and higher-governance asymptotic slope medians were 1.121 and 0.511, respectively, and the higher-minus-lower contrast was -0.601 (95% credible interval: -1.206 to -0.116), with $P(\text{higher} < \text{lower}) = 0.992$. A matched latent quadratic model (M3Q) slightly outperforms M4R in the audited three-model observation-level Pareto-smoothed importance sampling leave-one-out cross-validation (PSIS-LOO) comparison, indicating stronger evidence for concavity than for a uniquely threshold-shaped functional form. The final same-baseline validation gives M4R only a modest observation-level advantage over the robust latent linear model (M3R) and no aggregate advantage in the five-fold country-group validation. Threshold support and six source-audited prior sensitivity re-estimations preserved the negative asymptotic slope contrast, but not a precise cutoff. Therefore, the evidence supports weak structural threshold survival: a smaller higher-

governance asymptotic slope is credible, whereas a universal governance cutoff, monotonic decline in the full pointwise marginal derivative, predictive transportability, and causal effects remain unestablished.

Keywords: governance; measurement error; threshold model; Bayesian dynamic panel; Worldwide Governance Indicators; economic growth

JEL classification: C11; C23; O43

1. Introduction

Governance thresholds are policy-relevant because they can be interpreted as reform boundaries. Institutional economics links credible constraints, enforceable rights, predictable regulations, and administrative capacity to lower transaction costs and stronger investment incentives (Acemoglu et al., 2001; North, 1990). Yet a threshold is informative only if the governance score defining regime membership is measured precisely enough to support that classification.

Dynamic panel studies generally associate stronger institutions with better economic performance, while the effects vary across governance dimensions, regions, and income groups (Karagiannis & Paleologou, 2025; Misi Lopes et al., 2023; Ogbuabor et al., 2025; Parsa & Datta, 2023). Threshold studies allow regime-dependent returns but typically treat the governance variable as measured without error (Asamoah, 2021; Glawe, 2025; Ochi et al., 2023, 2024; Razzaq et al., 2023). This leaves a central gap: nonlinear flexibility does not resolve uncertainty in the variable assigning regimes.

WGI country-year estimates are latent-variable measures accompanied by producer-reported standard error (Kaufmann & Kraay, 2024). Heterogeneous measurement errors can weaken within-country signals, attenuate regime-specific slopes unequally, and reclassify observations near the

candidate cutoff. The dependence among component errors can further alter the composite scores (Griliches & Hausman, 1986; Magnusson & Tarverdi, 2020; Qu et al., 2019). The relevant question is therefore whether regime differentiation survives when uncertainty in the regime-defining variable is propagated through the model.

The analysis asks three questions: whether WGI uncertainty changes threshold location or regime classification; whether lower- and higher-governance slopes differ after joint latent measurement; and whether the higher-governance slope is smaller, as predicted by saturation, or larger, as predicted by complementarity. These mechanisms motivate observable slope patterns but are not directly measured causal channels.

The empirical framework uses 3,485 observations from 167 economies during 2002–2023. A Bayesian errors-in-variables dynamic panel model jointly estimates a persistent common governance factor across all six WGI dimensions, along with their reported standard errors, lagged growth dynamics, correlated country effects, centered year effects, lagged macroeconomic controls, and a smooth transition between lower- and higher-governance slopes. The Mundlak-type country-effect specification permits a correlation between persistent country heterogeneity and observed regressors, while the Student-t growth likelihood limits the leverage of crisis and rebound observations without removing them from the baseline sample. The framework addresses measurement uncertainty, dynamic persistence, observed macroeconomic confounding, and important forms of unobserved heterogeneity, but does not supply exogenous institutional variation. Therefore, the estimated governance parameters are conditional dynamic associations rather than causal effects. Empirically, the conventional cutoff is highly unstable under country bootstrap and composite measurement-error perturbations; however, the final audited latent model retains a credibly negative higher-minus-lower slope contrast. The cutoff posterior remains broad,

a latent quadratic specification is predictively competitive, and country-grouped validation does not establish an aggregate transport advantage for the threshold model.

The study contributes by separating nonlinear slope ordering from cutoff precision, integrating producer-reported WGI uncertainty with six-indicator latent measurement and dynamic growth in one audited Bayesian architecture, and distinguishing structural from predictive survival through resampling, measurement, functional-form, support, prior, temporal, and predictive checks. The policy implication is reporting discipline: estimated institutional thresholds should be presented as uncertain transition regions with probabilistic membership rather than deterministic universal targets.

The remainder of this paper is organized as follows. Section 2 integrates the institutional, nonlinear, and measurement-error literature and develops the hypotheses. Section 3 describes the sample, variable construction, and data handling rules. Sections 4 and 5 present the econometric framework, identification assumptions, computational diagnostics and threshold survival criteria. Section 6 reports the empirical results and hypothesis testing. Section 7 interprets the theoretical, methodological, and policy implications and evaluates the competing explanations and limitations. Section 8 concludes.

2. Literature Review and Hypothesis Development

2.1 Institutional quality and economic performance

Institutional economics treats governance as productive infrastructure. Credible constraints, enforceable property and contract rights, predictable regulations, and capable administration can lower transaction costs and protect investment returns (Acemoglu et al., 2001; North, 1990). These

mechanisms imply a positive baseline association with performance, but not necessarily constant marginal returns across the governance distribution.

Competing mechanisms imply different nonlinear shapes. Complementarity predicts larger returns when institutional capacities reinforce one another; credibility emphasizes durable, broad reforms; adjustment costs allow short-run disruption; and saturation predicts smaller annual-growth gains once foundational constraints are no longer binding. The empirical question is which mechanism best describes the observed governance distribution.

Recent evidence supports heterogeneous governance-growth relationships across income groups, regions, and WGI dimensions (Karagiannis & Paleologou, 2025; Misi Lopes et al., 2023; Ogbuabor et al., 2025; Parsa & Datta, 2023). This heterogeneity motivates nonlinear modeling but cautions against treating a single estimated threshold as an invariant institutional law.

Reverse causality and time-varying confounding remain important because growth can expand fiscal and administrative capacity while crises can weaken both growth and governance. Dynamic panel generalized method of moments (GMM) and Bayesian dynamic panel models address different parts of this problem, but neither creates exogenous institutional variation (Arellano & Bond, 1991; Baltagi et al., 2022; Blundell & Bond, 1998; Hirano, 2002). Accordingly, the estimated mechanisms are interpreted as conditional associations rather than causal effects.

2.2 Why governance effects may be nonlinear

A linear governance-growth model imposes a constant marginal association, whereas complementarity, adjustment, credibility, and saturation imply state dependence. In the smooth-transition model, complementarity predicts a larger higher-governance slope, saturation a smaller

one, and homogeneous returns little difference. The slope ordering is therefore the main empirical discriminator; transition location and speed describe where and how gradually the change occurs.

Threshold methods formalize regime dependence through fixed effects, dynamic, endogenous, threshold, and covariate-dependent extensions (Hansen, 1999; Kremer et al., 2013; Seo & Shin, 2016; Yang et al., 2024). Applied studies have reported nonlinear governance relationships across growth and related outcomes (Asamoah, 2021; Glawe, 2025; Ochi et al., 2023, 2024; Razzaq et al., 2023; Zhao et al., 2023). These findings make nonlinearity plausible without establishing a stable, portable policy cutoff.

The study organizes prior work along two dimensions: functional form (linear versus nonlinear) and measurement (exact versus uncertainty-modeled governance). Conventional regressions largely occupy the exact/linear cell, most threshold applications the exact/nonlinear cell, and uncertainty-adjusted index models often remain linear. The preferred design occupies the nonlinear/uncertainty-modeled cell and is compared directly with a latent-linear benchmark (Table 1).

Policy usefulness and statistical identification must be separated. A single cutoff is easy to communicate, but a sharp in-sample objective can coexist with location instability under resampling, measurement perturbation, support changes, or competing nonlinear forms. Credible threshold evidence therefore requires separate assessments of regime differentiation, cutoff concentration, boundary behavior, model dependence, and predictive performance.

2.3 Measurement uncertainty in the Worldwide Governance Indicators

The WGI reports six governance dimensions and an uncertainty estimate for each country-year-dimension observation. Because these scores are constructed from multiple surveys and expert

assessments, the reported standard errors are part of the measurement system, rather than ancillary statistics (Kaufmann & Kraay, 2024). Treating the point estimates as known with certainty discards this information.

Measurement precision varies across countries, years, and dimensions as source coverage and agreement change. Dependence assumptions can materially alter governance estimates, and corruption indicator standard errors vary systematically (Magnusson & Tarverdi, 2020; Qu et al., 2019). In panels, such heterogeneous reliability can make observed within-country changes combine true institutional movement with transitory measurement noise.

Classical covariate error attenuates slopes, especially when the within-country signal is weak relative to transitory noise (Griliches & Hausman, 1986). Regime models add misclassification: measurement error changes both the regressor and the probability of falling below or above the transition. Near a cutoff, this can create asymmetric attenuation and false location precision.

Because an uncertain institutional score determines regime membership, deterministic statements that an economy has crossed a reform boundary can overstate evidence. Decision problems with measurement uncertainty require explicit treatment of classification errors and structural uncertainty (Ferrero & Petri, 2025; Zamora & Towse, 2023). Posterior transition distributions, regime probabilities, and classification entropy are therefore more defensible reporting objects.

2.4 Bayesian errors-in-variables inference

Bayesian errors-in-variables models jointly estimate latent covariates, measurement parameters, dynamic states, and outcome parameters, allowing WGI uncertainty to propagate to the slopes, transition locations, and regime probabilities (Levy & McNeish, 2025; Sinha, 2021; Skarstein et

al., 2023). Joint estimation avoids treating an estimated institutional index as fixed in a second-stage nonlinear model.

Dynamic cross-country panels still require scale identification, separation of persistent institutional states from country heterogeneity and common shocks, and robust treatment of crisis observations. A hierarchical Bayesian formulation coherently accommodates these features (Baltagi et al., 2022; Hirano, 2002), but uncertainty propagation does not eliminate reverse causality or omitted-variable bias without exogenous institutional variation.

The integrated framework implies that measurement uncertainty may shift the cutoff location or regime classification; a nonzero lower-versus-higher asymptotic slope contrast indicates regime differentiation; and a negative higher-minus-lower contrast supports a smaller higher-governance asymptotic slope. Because the implemented governance contribution is a smoothly varying coefficient multiplied by latent governance, this contrast does not establish that the full pointwise marginal derivative decreases monotonically throughout the transition region. Threshold survival additionally requires an interpretable interior location and predictive value relative to the matched latent linear benchmark.

Table 1 Conceptual typology of governance-growth models

Outcome form	Governance is treated as exact	Governance uncertainty modeled
Linear	Conventional observed-score growth model (M0)	Latent measurement benchmark (M3R)
Nonlinear	Conventional threshold model (M1)	Composite EIV approximation (M2); latent nonlinear models (M3Q/M4R)

Notes: The typology separates assumptions regarding the functional form from assumptions regarding the measurement. Conventional governance-growth regressions occupy the exact-measurement/linear cell; most governance-threshold applications occupy the exact-measurement/nonlinear cell; and uncertainty-adjusted index studies typically occupy the uncertainty-modeled/linear cell. The preferred design evaluates a comparatively underdeveloped nonlinear/uncertainty-modeled combination against a latent linear benchmark.

2.5 Hypotheses

2.5.1 Measurement sensitivity

Threshold classification is measurement-sensitive because the same uncertain score determines both regressor value and regime assignment. Heterogeneous WGI standard errors and dependence among component errors can shift the inferred location or make posterior regime membership less decisive.

A competing view is that WGI uncertainty is small relative to cross-country dispersion and therefore immaterial. That argument is less persuasive near a regime boundary, where moderate score uncertainty can reverse classification. H1 therefore concerns location and membership sensitivity rather than mechanical attenuation of every coefficient.

H1: Incorporating WGI measurement uncertainty materially changes the threshold location and/or regime classification relative to an analysis that treats the governance score as error-free.

2.5.2 Regime differentiation

Institutional theories allow governance returns to vary across the institutional distribution because legal enforcement, administrative implementation, regulatory credibility, and corruption control can relax different constraints in low- and high-capacity settings. A constant slope can therefore mask economically meaningful state dependence.

The competing homogeneous-return mechanism predicts that apparent regime differences disappear once measurement error, persistence, and country heterogeneity are modeled. Evidence for H2 consequently requires a credible posterior slope contrast within the identified latent measurement system.

H2: The lower- and higher-governance asymptotic slope parameters differ within the latent smooth transition model.

2.5.3 Diminishing marginal returns

The saturation mechanism predicts smaller annual-growth associations at higher governance because foundational institutional constraints are less binding, even though further improvements may remain valuable for resilience, rights, distribution, and long-run welfare.

Complementarity predicts the opposite ordering: institutional reforms may become more productive after complementary state capacities are established. The empirical distinction therefore depends on the posterior comparison of the two asymptotic slopes, not on the cutoff estimate alone.

H3: The higher-governance asymptotic slope parameter is smaller than that of the lower-governance asymptotic slope parameter.

2.5.4 Threshold-survival evaluation rule

Threshold survival is an ex-ante reporting rule rather than a fourth hypothesis. Structural survival requires credible regime separation, an interpretable interior threshold, and qualitative stability across measurement, support, and prior assumptions; predictive survival additionally requires incremental information relative to the latent linear benchmark in observation-level and country-group validation.

3. Data and Variable Construction

3.1 Sample

The empirical sample merges the Worldwide Governance Indicators with the World Development Indicators for 2002-2023. The start year marks the transition to consistent annual WGI coverage, and the endpoint ensures a common data vintage across all six governance estimates, their reported standard errors, growth outcomes, and baseline controls (Kaufmann & Kraay, 2024). The unit of analysis is the economy-year. Applying the common support rules described below yields 3,485 observations from 167 economies.

Sample construction follows a prespecified common support procedure designed to keep the measurement and outcome systems aligned. Aggregate regions and non-economy entities are removed, country identifiers are harmonized across sources, and an economy-year is retained only when real GDP per capita, all six WGI point estimates and standard errors, and the baseline controls are jointly observed. Each economy contributed at least eight consecutive complete observations. The requirement preserves sufficient temporal information for the latent state process while avoiding the much larger coverage loss that a fully balanced global panel would impose.

The resulting panel is unbalanced but dense: country-specific histories range from 8 to 22 years, the median is 22 years, and 139 economies contribute to all 22 annual observations. The baseline does not interpolate, forward-fill, or model impute missing inputs. This choice makes the estimation sample transparent but does not remove the selection induced by the common support. Therefore, the reported estimates describe economies that satisfy the stated coverage criteria, and the consecutive coverage sensitivity analysis evaluates how strongly the main nonlinear inference depends on that selection rule.

3.2 Dependent variable

The dependent variable is annual growth in real GDP per capita (g_{it}). Growth is defined as 100 times the first difference of the natural logarithm of constant-price real GDP per capita (Y_{it}), a transformation that places economies on a comparable percentage scale and focuses the analysis on short-run performance rather than a potentially nonstationary income level:

$$g_{it} = 100[\ln(Y_{it}) - \ln(Y_{i,t-1})],$$

The economy index (i) and year index (t) identify each country-year observation. Real GDP per capita is the World Development Indicators series NY.GDP.PCAP.KD and is measured in constant 2015 US dollar. Lagged growth captures short-run persistence and reduces the omitted dynamic bias. Extreme annual observations are retained in the preferred specification because crises and rebounds are economically meaningful; their leverage is limited through a student-t likelihood. Robustness checks assess annual growth winsorized at the 1st and 99th percentiles, the exclusion of absolute growth rates above 30%, and a three-year forward average defined as the mean of annual growth one, two, and three years ahead to determine whether the threshold pattern depends on transitory volatility or outcome timing.

3.3 Governance indicators and reported uncertainty

The governance-dimension index j runs from 1 to 6. Let q_{it} denote latent institutional quality for economy i in year t ; z_{jit} is the published WGI point estimate for dimension j , and s_{jit} is its producer-reported standard error. In the 2024 reproducibility package, WGI point estimates were approximately scaled from -2.5 to +2.5, with higher values indicating stronger governance. Table 2 lists the six dimensions, codes and interpretations.

All six WGI dimensions enter with observation-specific standard errors, allowing the measurement precision to vary across country-year-dimension observations. For M0-M2 only, the observed equal-weight composite is $\bar{g}_{it} = (1/6)\sum_{j=1}^6 z_{jit}$. Its aggregated standard error s_{it} is $(1/6)\sqrt{(\sum_{j=1}^6 s_{jit}^2)}$ under independent component errors and $(1/6)\sum_{j=1}^6 s_{jit}$ under perfect positive correlation. The preferred latent model jointly estimates the six dimensions and allows for residual dimension-specific dispersion beyond the reported WGI standard errors.

Table 2 Worldwide Governance Indicators dimensions and institutional interpretations

Code	Governance dimension	Institutional interpretation
VA	Voice and Accountability	Participation, political rights, media freedom, and accountability
PV	Political Stability and Absence of Violence/Terrorism	Stability of political authority and exposure to politically motivated violence
GE	Government Effectiveness	Quality of public services, civil service, policy formulation, and implementation credibility
RQ	Regulatory Quality	Capacity to formulate and implement policies supportive of private-sector development
RL	Rule of Law	Contract enforcement, property rights, courts, police, and protection from crime
CC	Control of Corruption	Constraints on the use of public power for private gain

The preferred M3R/M4R measurement system treats institutional quality as a latent common factor. Each WGI dimension has a dimension-specific intercept, loading, and residual dispersion. Rule of Law anchors factor location and scale through an intercept fixed at zero and a loading

fixed at one; the remaining loadings use weakly regularizing Normal(1, 0.5) priors without a hard positivity constraint. The residual component captures dispersion beyond the producer-reported World Governance Indicator (WGI) standard errors.

3.4 Control variables

The one-period-lagged control vector ($\mathbf{x}_{i,t-1}$) comprises lagged log real GDP per capita, gross capital formation as a share of GDP, trade openness, the logarithm of one plus the absolute inflation rate, and population growth. The variables follow standard growth empirics and recent dynamic governance applications: initial income represents conditional convergence; investment captures capital accumulation; trade openness represents market access and technology diffusion; inflation proxies macroeconomic instability; and population growth captures capital dilution and labor-supply dynamics (Acemoglu et al., 2001; Parsa & Datta, 2023). Continuous controls are standardized within the estimation sample, and the economy-specific means of the standardized control vector ($\bar{\mathbf{x}}_i$) enter the Bayesian models through a Mundlak correlated-random-effects specification (Mundlak, 1978). Table 3 summarizes the operational definitions, original units, transformations, lag structure, standardization, and roles in the growth equation. Because investment and trade can also transmit institutional effects, the baseline coefficient is interpreted as a conditional governance-growth association, holding these observed macroeconomic conditions constant rather than as the total association operating through all transmission channels.

Table 3 Baseline control variables, operational definitions, and roles in the growth equation

Variable	Operational definition, unit, and transformation	Role in the growth equation
Initial income	One-period lag of $\ln(\text{real GDP per capita, constant 2015 US\$})$; standardized within the estimation sample	Conditional convergence
Investment	One-period lag of gross capital formation (% of GDP); standardized within the estimation sample	Capital accumulation
Trade openness	One-period lag of exports plus imports (% of GDP); standardized within the estimation sample	Market access and technology diffusion
Inflation	One-period lag of $\ln[1 + \text{GDP-deflator inflation rate}]$; standardized within the estimation sample	Macroeconomic instability
Population growth	One-period lag of annual population growth (%); standardized within the estimation sample	Capital dilution and labor-supply dynamics
Country means	Economy-specific means of the standardized control vector	Mundlak correlated effects

Notes: Each variable listed in Table 3 enters the growth equation through a lagged control vector. All underlying series were obtained from the World Development Indicators. Percentage variables are recorded in percentage points prior to within-sample standardization, and each baseline control enters with a one-year lag effect. The economy-specific means of the standardized control vector are included in the correlated random effects component.

Human capital is excluded because available series materially reduce global common support and may also lie on a governance-to-growth pathway. Investment and trade are retained, so the baseline estimates a governance-growth association conditional on major macroeconomic channels rather than the total association through all transmission mechanisms. This trade-off prioritizes a common auditable sample and is treated as an interpretive limitation.

3.5 Descriptive evidence

Descriptive evidence confirms substantial outcome heterogeneity and nonconstant measurement precision. Annual real GDP-per-capita growth averages 1.888%, with a standard deviation (SD) of 5.545%, and ranges from -80.505% to 65.119%, indicating crisis and rebound observations that would be highly influential under a thin-tailed distribution. The equal-weight WGI composite had a mean of 0.002 and an SD of 0.899, with a range from -2.221 to 1.947. Under independent component errors: mean composite standard error 0.083; SD 0.019; range 0.064–0.276. Figure 1 shows that the average reported WGI uncertainty declined through the early 2010s and rose again toward the end of the sample. The heterogeneity across time and country-years supports an observation-specific measurement likelihood rather than a constant-reliability adjustment.



Figure 1 Average reported WGI standard error by year

Notes: The figure averages the six producer-reported WGI standard errors across the available country-year observations.

3.6 Data handling and pre-estimation assessment

Common support and consecutive-coverage rules are applied before estimation. The final input has no missing growth, lagged growth, controls, WGI estimates, or WGI standard errors; no interpolation or imputation is used. Extreme growth observations remain in the primary sample and are handled with a Student-t likelihood, while winsorization and exclusions are robustness checks. Alternative history requirements assess selection induced by common support.

The estimand is a conditional short-run growth association rather than a long-run cointegrating relation. Growth is modeled in log differences and the latent institutional state is stationary, so panel cointegration tests are not model-selection gates. Centered year effects, correlated country effects, and pandemic-year sensitivity address common shocks and persistent heterogeneity without claiming that all structural breaks or time-varying confounding are removed.

3.7 Notation and reporting conventions

The substantive variables and panel indices are introduced in Sections 3.2 to 3.4. The indicator function ($I(\cdot)$) equals one when its stated condition is satisfied and zero otherwise.

In the measurement and state equations, a_j denotes the dimension-specific intercept, b_j the factor loading, and τ_j the residual measurement standard deviation. The terms μ_i and δ_t denote the economy-specific and common-year components of latent institutional quality, respectively; φ is the latent institutional persistence coefficient; η_{it} is the state innovation; and σ_q is its innovation standard deviation, so σ_q^2 is the innovation variance.

In the growth equation, the country-specific effect (α_i), common year effect (λ_t), lagged-growth coefficient (ρ), lower-governance slope (β_L), higher-governance slope (β_H), threshold location (γ), transition-speed parameter (κ), control-coefficient vector (θ), and outcome

disturbance (ε_{it}) enter the conditional outcome specification. The smooth-transition weight (w_{it}), governance-specific slope coefficient ($\beta(q_{it})$), remaining outcome-parameter set (Ω), Student-t degrees-of-freedom parameter (ν), and growth-likelihood scale parameter (σ_g) complete the outcome model. Here, Ω collectively denotes the remaining outcome-hierarchy parameters not separately conditioned on in the displayed likelihood, including country and year effects, lagged-growth and control terms, slope/transition terms, Student-t degrees of freedom, and the growth scale.

In the correlated-random-effects component, the country-effect intercept α_0 , economy-specific standardized-control means $\bar{x}_i$, coefficient vector on those means π , economy-specific latent-governance mean $\bar{q}_i$, coefficient on mean latent governance π_q , and country-effect variance σ_α^2 define the hierarchical country effect. For posterior classification, the higher-governance regime probability p_{it}^H , regime-classification entropy H_{it} , observed data D , and regime-slope contrast $\Delta_\beta = \beta_H - \beta_L$ enter the threshold-survival criteria. The statistical separation is defined in Section 5.3. The magnitude threshold $\delta_{\min}$ equals 1 for the stronger one-point criterion. It is measured on the same higher-minus-lower growth-slope scale as Δ_β . Because annual growth is defined as 100 times a log difference, the lower- and higher-governance slopes, their contrast, and the magnitude threshold are expressed in annual 100-log-point growth units per unit of rule-of-law-anchored latent governance.

4. Econometric Framework

4.1 Conventional dynamic panel threshold model

A conventional one-threshold growth model can be written as

$$g_{it} = \alpha_i + \lambda_t + \rho g_{i,t-1} + \beta_L q_{it} I(q_{it} \leq \gamma) + \beta_H q_{it} I(q_{it} > \gamma) + x'_{i,t-1} \theta + \varepsilon_{it}$$

All variables and parameters in the conventional threshold equation are defined in Sections 3.2–3.7. Conventional applications replace latent institutional quality with an observed WGI composite and classify regimes using point estimates.

The preceding threshold specification becomes an errors-in-variables problem because q_{it} is not observed directly. A simple representation is

$$\tilde{q}_{it} = q_{it} + u_{it}, \quad u_{it} \sim \mathcal{N}(0, s_{it}^2),$$

The published or constructed governance score ($\tilde{q}_{it}$), measurement error (u_{it}), and aggregated score uncertainty (s_{it}) define the composite approximation. Replacing latent institutional quality with the observed score affects both regressors and regime indicators.

4.2 Hierarchical WGI measurement model

The preferred specification directly uses the six WGI dimensions. Let z_{jit} be the published estimate and s_{jit} its reported standard error. The measurement model is

$$z_{jit} \mid q_{it}, a_j, b_j, \tau_j \sim \mathcal{N}(a_j + b_j q_{it}, s_{jit}^2 + \tau_j^2), \quad j = 1, \dots, 6.$$

The observation-specific measurement variance combines the producer-reported WGI standard error with the residual dimension-specific dispersion. Therefore, the likelihood discounts less precise indicator observations while allowing the data to reveal additional disagreement among dimensions. This approach treats the WGI uncertainty estimate as part of the measurement process rather than as a post-estimation weight or an error-free input.

Factor identification uses the Rule of Law intercept fixed at zero and loading fixed at one; remaining loadings are weakly regularized rather than sign-constrained. Centering country and

year components prevents location drift and separates persistent cross-country heterogeneity from common annual variation. The resulting latent factor is rule-of-law anchored, so raw coefficients under alternative anchors are not directly scale-comparable.

4.3 Latent institutional dynamics

Institutional quality follows a stationary, first-order autoregressive state process. Persistence is substantively necessary because legal and administrative capacity changes gradually, while stationarity prevents the latent path from absorbing unrestricted trends that belong to country or year effects.

$$q_{it} = \mu_i + \delta_t + \varphi(q_{i,t-1} - \mu_i) + \eta_{it}, \quad \eta_{it} \sim \mathcal{N}(0, \sigma_q^2),$$

Institutional persistence is constrained to (0,1) by the implemented mapping $\varphi = \text{inv_logit}(\varphi_{\text{raw}})$, with $\varphi_{\text{raw}} \sim \text{Normal}(1.07, 0.35)$; the initial state uses the corresponding stationary variance. Centered country components capture persistent differences, centered year components common shifts, and innovations country-year departures. Stationarity is a maintained modeling assumption rather than evidence that published WGI estimates are trend-free.

4.4 Bayesian errors-in-variables dynamic threshold outcome model

For comparison, substituting latent q_{it} into the conventional discontinuous threshold representation yields

$$g_{it} \mid q_{it}, \gamma, \Omega \sim t_v(\alpha_i + \lambda_t + \rho g_{i,t-1} + [\beta_L I(q_{it} \leq \gamma) + \beta_H I(q_{it} > \gamma)]q_{it} + x_{i,t-1}'\theta, \sigma_g)$$

The latent hard-threshold form is used only as a conceptual benchmark. Country effects follow a correlated random-effects structure and the growth disturbance a Student-t distribution, retaining

crises and rebounds while limiting Gaussian leverage. Reported M4R inference comes from the smooth-transition specification.

$$\alpha_i \sim \mathcal{N}(\alpha_0 + \bar{\mathbf{x}}_i' \boldsymbol{\pi} + \pi_q \bar{q}_i, \sigma_\alpha^2).$$

Mundlak allows persistent country heterogeneity to correlate with observed regressors (Mundlak, 1978). Growth parameters are therefore informed by within-economy changes, between-economy information conditional on those means, and common-year adjustment. This relaxes random-effects independence but does not create randomized or quasi-random governance variation.

M4R replaces the discrete regime assignment with differentiable logistic interpolation between the lower- and higher-governance asymptotic slope parameters. The implemented governance contribution is $\beta(q_{it})q_{it}$, where the coefficient varies smoothly with latent governance through the logistic transition weight. This parameterization permits the Hamiltonian Monte Carlo and allows the association to change over a range rather than at a literal discontinuity. Because this is only one nonlinear form, M4R is compared with matched latent quadratic and linear alternatives.

$$w_{it} = [1 + \exp\{-\kappa(q_{it} - \gamma)\}]^{-1},$$

$$\beta(q_{it}) = \beta_L(1 - w_{it}) + \beta_H w_{it},$$

$$g_{it} \mid q_{it}, \gamma, \kappa, \Omega \sim t_v(\alpha_i + \lambda_t + \rho g_{i,t-1} + \beta(q_{it}) q_{it} + \mathbf{x}'_{i,t-1} \boldsymbol{\theta}, \sigma_g)$$

The transition weight w_{it} is the model-implied logistic interpolation weight that determines the governance-specific coefficient in the growth equation. It is distinct from the posterior high-regime classification probability $p_{it}^H = \Pr(q_{it} > \gamma \mid D)$ defined in Section 5.2, which summarizes posterior uncertainty about whether latent governance exceeds the transition location. The parameter κ controls how rapidly w_{it} moves between 0 and 1. In the implemented level equation,

$\beta(q_{it})$ is a smoothly varying coefficient; β_L and β_H are the asymptotic lower- and higher-governance slope parameters. Within the transition region, the pointwise derivative of the level contribution also reflects the logistic weight derivative. The Student-t likelihood limits sensitivity to extreme annual growth.

4.5 Prior distributions

Priors are weakly regularizing and scaled to standardized covariates. The locked executable M4R uses Normal(0, 1) free WGI intercepts, Normal(1, 0.5) free loadings, a fixed Rule of Law anchor, stationary persistence, a positive Lognormal(log 8, 0.35) transition speed, and an interior-regularizing threshold prior. Table 4 reports the implemented specifications following the explicit Bayesian reporting practice (Kruschke, 2021).

Table 4 Prior distributions and parameter constraints

Parameter	Prior or constraint	Purpose
block		
WGI intercepts	Free $a_j \sim \text{Normal}(0, 1)$; Rule of Law intercept fixed at 0	Measurement location identification
WGI loadings	Free $b_j \sim \text{Normal}(1, 0.5)$; Rule of Law loading fixed at 1; no hard positivity constraint	Scale/sign orientation through the fixed anchor; weak regularization
Residual WGI SDs	$\tau_j > 0$; Normal(0, 0.15) on the positive scale	Uncertainty beyond reported WGI SEs

Parameter	Prior or constraint	Purpose
block		
Institutional persistence	$\varphi_{\text{raw}} \sim \text{Normal}(1.07, 0.35); \varphi = \text{inv_logit}(\varphi_{\text{raw}}), 0 < \varphi < 1$	Stationary positive latent-state persistence
Regime slopes	$\beta_L, \beta_H \sim \text{Normal}(0, 3)$	Weak regularization of governance effects
Threshold location	$\gamma_{\text{raw}} \sim \text{Beta}(2, 2); \gamma = \gamma_L + (\gamma_U - \gamma_L)\gamma_{\text{raw}},$ where the lower and upper symbols denote the admissible support bounds.	Interior threshold regularization
Transition speed	$\kappa > 0; \kappa \sim \text{Lognormal}(\log 8, 0.35)$	Smooth-to-sharp transition without a hard upper bound
Growth likelihood	$\sigma_g > 0; \sigma_g \sim \text{Normal}(0, 4)$ on the positive parameter scale; $v - 2 \sim \text{Exponential}(0.10)$	Heavy-tailed robust outcome model

Notes: SD denotes standard deviation and SE denotes standard error. Table 4 reflects the locked executable M4R source after the specification-integrity audit. The rule of law intercept and loading are fixed at 0 and 1, respectively; the remaining loadings are not hard-constrained to be positive. Distributional statements report the implemented priors or parameter constraints, rather than provisional development-stage settings.

Conditional prior-predictive simulation evaluates whether the prior architecture generates plausible growth behaviors before observing outcomes (Winter & Depaoli, 2023). Threshold-support sensitivity uses 5th–95th, 10th–90th, and 15th–85th percentile supports; source-audited prior variants alter slope regularization, measurement dispersion, persistence, and transition smoothness. These checks distinguish slope robustness from location dependence.

5. Identification, Computation, and Threshold-Survival Tests

5.1 Statistical identification under maintained assumptions

Identification is separated into measurement, conditional-outcome, and causal components. The latent WGI factor is statistically identified under the Rule of Law anchor, repeated indicators, observation-specific standard errors, stationarity, weak loading regularization, and centered country/year components. This identification is conditional on the anchor and common-factor assumptions.

Regime slopes and transition parameters use within-economy institutional variation, cross-economy differences conditional on Mundlak means, lagged growth, common-year effects, and observations spanning latent support. Because M3R, M3Q, and M4R share the same measurement, state, country-effect, control, and Student-t architecture, their predictive comparisons isolate functional-form differences.

The design does not causally identify an exogenous governance effect. No external instrument, natural experiment, policy discontinuity, or other plausibly exogenous institutional variation is used. Bayesian estimation propagates uncertainty but cannot by itself eliminate reverse causality or omitted-variable bias; governance slopes are therefore interpreted as conditional dynamic associations.

5.2 Posterior regime probabilities

Under measurement uncertainty, the regime membership is probabilistic. For each observation, define

$$p_{it}^H = \Pr(q_{it} > \gamma \mid \mathcal{D}),$$

Values near zero or one indicate a relatively decisive posterior regime classification, whereas values near one-half indicate substantial ambiguity. Regime probabilities are retained for each economy-year so that classification uncertainty can be reported directly rather than being replaced by deterministic statements about whether an economy has crossed the estimated transition.

A regime-classification entropy statistic is

$$H_{it} = -p_{it}^H \ln p_{it}^H - (1 - p_{it}^H) \ln(1 - p_{it}^H).$$

Higher entropy corresponds to greater posterior uncertainty regarding membership in lower- versus higher-governance regimes. Therefore, the entropy measure is a classification diagnostic, not an additional estimate of the magnitude of the economic effect.

5.3 Definition of threshold survival

Threshold survival is evaluated separately for structural and predictive evidence. Structural survival requires credible regime separation, an interior location, and qualitative stability across measurement, support, and prior assumptions. Predictive survival additionally requires incremental information over the latent linear benchmark under observation-level and country-group validation.

Condition 1 requires the 95% credible interval for the higher-minus-lower slope to exclude zero and $P(\text{higher} < \text{lower}) > 0.95$; a separate magnitude criterion prevents a credible but trivial contrast from being described as large.

$$\Pr(|\Delta_\beta| > \delta_{\min} \mid \mathcal{D}) > 0.95,$$

Condition 2 assesses transition location and boundary behavior. Posterior concentration in the lowest or highest 5% of admissible support signals possible boundary-driven identification;

credible-interval width and support sensitivity determine whether the location can be reported as a cutoff, a broad transition region, or only as slope ordering.

Condition 3 compares M4R with M3R using PSIS-LOO and five-fold country-grouped cross-validation (CV). The former assesses predictions for observations from represented economies; the grouped design withholds each economy's growth-likelihood contribution while retaining its observed histories, so it evaluates conditional country-level transportability rather than fully unconditional prediction (Adin et al., 2024; Cooper et al., 2024). Predictive differences are interpreted based on their uncertainty (Bates et al., 2024; Sivula et al., 2025).

Condition 4 requires qualitative stability of slope ordering across measurement-error scenarios, support restrictions, sample, timing, likelihood, and measurement perturbations. Full survival requires both structural and predictive criteria; weak structural survival allows stable slope ordering with an imprecise location, functional-form ambiguity, or limited predictive gain.

5.4 Model sequence

The model sequence progressively relaxes exact measurement and linearity: M0 is the observed-score linear benchmark, M1 the conventional sharp threshold, M2 a composite measurement-error approximation, M3R the robust latent linear benchmark, and M4R the preferred latent smooth transition. M3R uses a single linear term for the latent governance factor. M3Q is the matched quadratic comparator, which uses separate first- and second-order terms in the same latent governance factor under the same latent measurement and dynamic architecture (Table 5).

Table 5 Model sequence and empirical purposes

Model	Governance treatment	Outcome form	Purpose
M0	Equal-weight WGI composite treated as exact	Linear Gaussian fixed-effects benchmark	Conventional association
M1	Equal-weight WGI composite treated as exact	Grid-searched one-threshold model	Naïve cutoff benchmark
M2	Composite perturbed using aggregated WGI uncertainty	Simulation-based threshold approximation	Measurement-dependence sensitivity
M3R	Six-indicator latent WGI factor with reported SEs	Linear Student-t dynamic model	Robust latent linear benchmark
M4R	Six-indicator latent WGI factor with reported SEs	Smooth-threshold Student-t dynamic model	Preferred nonlinear model

Notes: The suffix R denotes the robust Student-t outcome specification. M0-M2 use observed or perturbed composite governance measures, whereas M3R, M3Q, and M4R estimate governance through the six-indicator latent WGI measurement system with producer-reported standard errors. M3Q uses first- and second-order terms in the same latent WGI factor as the matched quadratic functional form benchmark.

M3R, M3Q, and M4R share the latent measurement system, state process, correlated country effects, controls, and heavy-tailed likelihood; therefore, their predictive differences isolate the governance functional form rather than a different treatment of measurement uncertainty or outcome tails.

5.5 Computation and diagnostics

All Bayesian models were implemented in Stan via CmdStanPy. The final audited M3R, M3Q, and M4R comparison uses four chains and 8,000 retained draws per model. The primary M4R fit uses the 3,485-observation sample, Rule of Law anchor, 5th–95th threshold support, and four chains with 2,000 warmup and 4,000 retained iterations each. It passed the strict gate: maximum key R-hat 1.0026, minimum key bulk effective sample size 1,698.3, zero divergences, and zero maximum tree depth hits.

Parameter recovery uses 50 replications under strong, weak, and null regime separation. The canonical set retained 123 primary fits, 22 anchor-tau initialization rescues, and 5 prior-centered latent-state rescues; rescue changed initial values only. Recovery is stronger for the regime contrast than for latent-state persistence, and null-threshold coverage is not interpreted as cutoff identification. No final audited rolling-origin result was available.

5.6 Executed robustness checks

Robustness checks target distinct threats: M3R/M3Q/M4R address functional form; anchors, WGI-SE scaling, and dimension omission address measurement; income-ranked-group-by-year effects common shocks; a two-piece Student-t likelihood tail asymmetry; coverage rules sample selection; governance timing temporal alignment; and support/prior re-estimation of posterior dependence. Final same-baseline PSIS-LOO and country-grouped validation address prediction. In Step 16E, the four income-ranked groups are country-level quartiles of the mean standardized lagged log real GDP per capita in the analysis sample; they are not official World Bank income groups. The group-by-year effects are constrained to sum to zero within each year and augment the common-year effect.

Conventional threshold stability is examined separately because the observed-score cutoff differs conceptually from the latent transition. Country bootstrap, trimming, tail exclusions, microstate and pandemic exclusions, lagged governance, forward growth, fixed-date interactions, and an exploratory candidate-year scan evaluate sampling, composition, timing, and temporal constancy without reinterpreting these estimates as M4R thresholds.

Measurement robustness proceeds from M2 composite perturbations under independent and perfectly correlated component errors to M3R/M4R joint latent measurement. This separates sensitivity caused by perturbing a summary score from uncertainty propagated jointly through measurement, institutional dynamics, and the growth equation.

Prior predictive checks, support re-estimation, source-audited prior variants, final PSIS-LOO, country-grouped validation, and a locked-posterior predictive check address distinct inferential risks. Development-stage posterior-predictive files remain only in provenance. The submission-level posterior predictive check (PPC) uses the archived `growth_rep` generated quantities from all 16,000 retained draws in the four exact locked Final Audited M4R chains; no model refit, Stan recompilation, or replacement predictive simulation is involved. The PPC evaluates in-sample generative adequacy and is conceptually distinct from PSIS-LOO and country-grouped transport validations. Therefore, robustness is reported separately for asymptotic slope stability, cutoff precision, posterior predictive adequacy, predictive transportability, prior dependence, and posterior reliability.

6. Empirical Results

6.1 Sample structure and conventional benchmarks

The results are organized to distinguish conventional apparent precision from the uncertainty-adjusted nonlinear evidence. The estimation sample contains 3,485 economy-year observations from 167 economies during 2002–2023. Country histories range from 8 to 22 years, with a median of 22 years, and 139 economies span the full 22-year range. Table 6 reports the sample structure and descriptive statistics that provide empirical support for the models.

Table 6 Sample structure and selected descriptive statistics

Measure	Value
Country-year observations	3,485
Economies	167
Period	2002–2023
Country-specific T	Minimum 8; median 22; maximum 22
Economies observed for all 22 years	139 (83.2%)
Real GDP per capita growth	Mean 1.888; SD 5.545; range -80.505 to 65.119
Equal-weight WGI composite	Mean 0.002; SD 0.899; range -2.221 to 1.947
Composite SE under independent errors	Mean 0.083; SD 0.019; range 0.064 to 0.276

Notes: T denotes the number of annual observations contributed by an economy; SD denotes the standard deviation; SE denotes the standard error; and WGI denotes the Worldwide Governance Indicators.

The observed score benchmarks initially produce a conventional governance-growth pattern that appears both positive and nonlinear. M0 estimates a positive linear governance coefficient of 4.698 (95% confidence interval: 2.542–6.855). M1 locates a naïve threshold at 0.847, with a lower regime slope of 5.676 (95% confidence interval: 3.108 to 8.245) and a higher regime slope of 2.157 (-0.550 to 4.864); the higher-minus-lower difference is -3.520 ($p=0.043$). Therefore, the

benchmark points toward a smaller higher-governance regime slope rather than a sign reversal, but it conditions on an error-free composite score and should not be interpreted as the preferred institutional threshold.

Sampling uncertainty substantially weakens the apparent precision of the M1 cut-off. Across the sample and outcome perturbations, the estimated thresholds ranged from 0.442 to 0.976. More importantly, the 2,000-replication country bootstrap yields a median of 0.816 but a 95% percentile interval from -0.922 to 1.026; only 44.0% of bootstrap thresholds lie within 0.10 of the naïve value of 0.847. Figure 2 illustrates an important distinction: the in-sample residual sum of squares (RSS) profile can exhibit a sharp minimum, even when the estimated location is unstable under resampling. The conventional model motivates nonlinear analysis but does not justify a universal cutoff value.

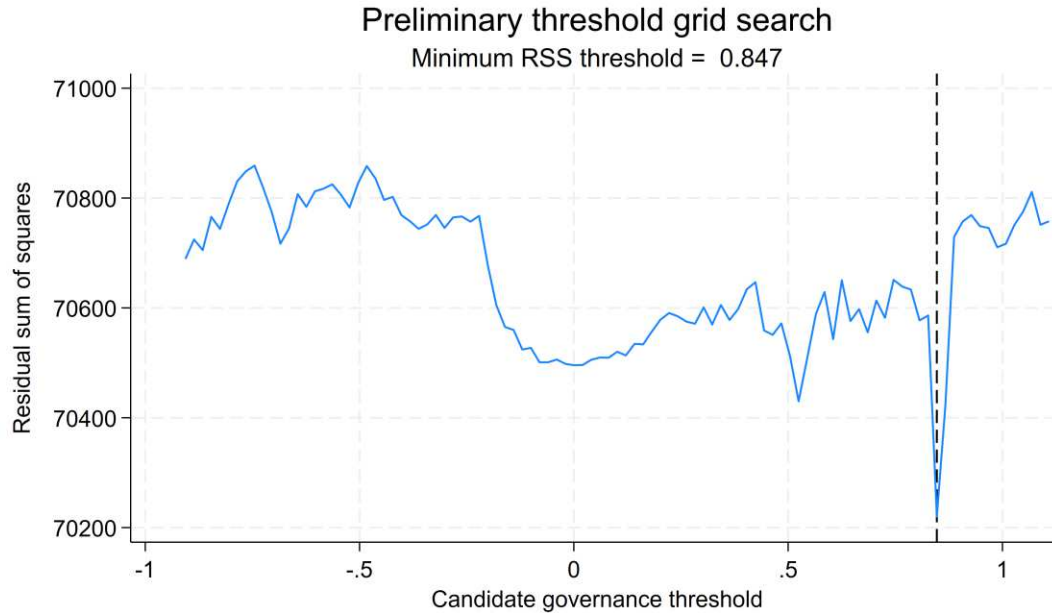


Figure 2 Naïve threshold grid search

Note: The dashed line marks the minimum RSS threshold of 0.847 on the equal-weight WGI composite.

6.2 Composite errors-in-variables approximation

M2 evaluates H1 by propagating the composite WGI uncertainty prior to the threshold estimation. Under independent component errors, the threshold median was 0.618, with a 95% Monte Carlo interval of -0.063 to 1.000; the mean lower- and higher-regime slopes were 4.646 and 1.206, respectively. Under a perfect positive correlation, the median threshold moved to 0.465 with a much broader interval of -0.943 to 1.079, and the corresponding slope means were 2.451 and 0.322. The higher-minus-lower differences remain negative under both dependence assumptions, but the location of the transition is materially less stable than the naïve estimate.

The contrast between the two M2 scenarios clarifies the substantive content of H1. The dependence structure of measurement errors changes both the center and dispersion of the inferred cutoff, while the probability of a negative slope ordering remains essentially one in the perturbation exercise. Therefore, Figure 3 and Table 7 indicate that propagating measurement uncertainty primarily reduces confidence in deterministic regime classification rather than eliminating the broader negative slope-ordering pattern observed in the results. H1 is supported because the treatment of WGI uncertainty materially changes the threshold inference relative to the error-free composite benchmark.

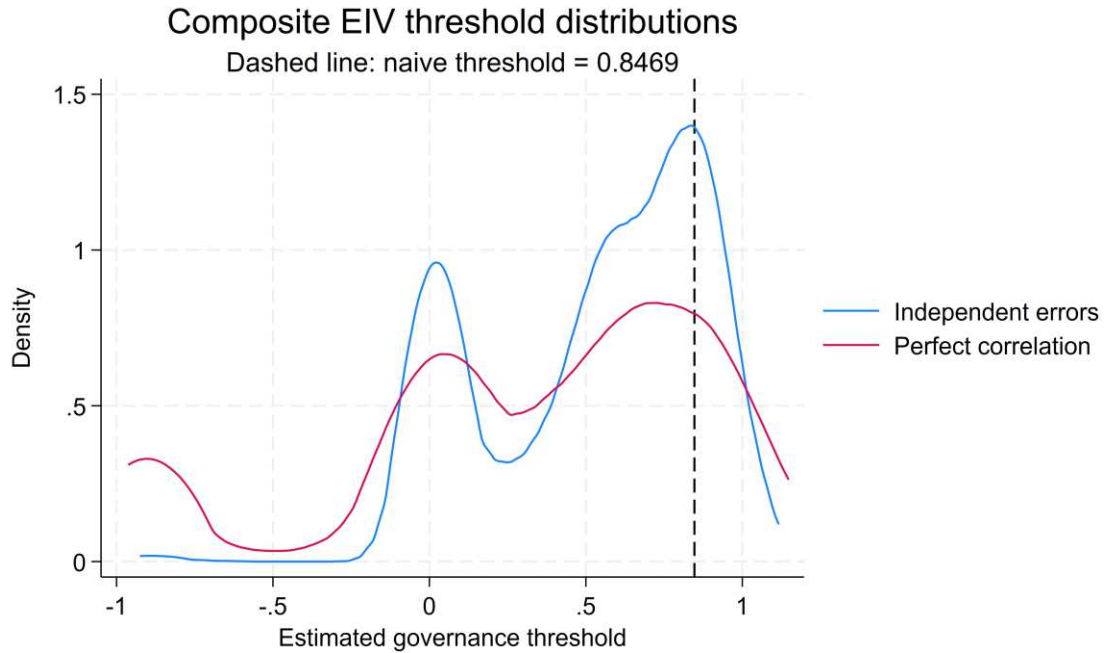


Figure 3 Composite errors-in-variables threshold distributions

Note: The dashed line represents the naïve threshold. The curves summarize 2,000 retained perturbation draws under alternative dependence assumptions for the six WGI component errors. EIV denotes errors in variables.

Table 7 Conventional and composite errors-in-variables threshold estimates

Model/scenario	Threshold	Lower slope	Higher slope	Higher–lower	Uncertainty statement
M1 naïve	0.847	5.676	2.157	-3.520	Difference p=0.043
M2 independent errors	Median 0.618	Mean 4.646	Mean 1.206	Mean -3.440	95% Monte Carlo percentile interval: -0.063 to 1.000

Model/scenario	Threshold	Lower slope	Higher slope	Higher–lower	Uncertainty statement
M2 perfect correlation	Median 0.465	Mean 2.451	Mean 0.322	Mean -2.129	95% Monte Carlo percentile interval: -0.943 to 1.079

Notes: M1 is the naïve observed-score threshold model and M2 is the composite errors-in-variables approximation. The M1 p-value is the two-sided test of equality between the higher and lower regime slopes. The M2 uncertainty statements are 95% Monte Carlo percentile intervals from 2,000 perturbation draws.

6.3 Latent measurement and linear dynamic model

The M3R measurement system provides the final audited latent linear comparator before the introduction of nonlinear slope differentiation. With the Rule of Law fixed at one, the posterior mean loadings are 0.803 for Voice and Accountability, 0.729 for Political Stability, 0.978 for Government Effectiveness, 0.942 for Regulatory Quality, and 0.995 for Control of Corruption. Residual measurement dispersion is larger for Voice and Accountability and Political Stability than for Rule of Law, indicating that producer-reported standard errors contain relevant information but do not exhaust cross-dimension measurement disagreement.

The latent institutional process is highly persistent: the posterior mean of the autoregressive coefficient is 0.977, with a 90% credible interval of 0.970–0.983. The latent linear governance slope was positive at 0.803 (90% credible interval: 0.269–1.417), and the lagged growth coefficient was 0.345 (90% credible interval: 0.318–0.372). The institutional innovation standard deviation was 0.057 (90% credible interval: 0.054 to 0.060), and the Student-t degrees-of-freedom parameter remained close to two. M3R therefore supplies the final audited linear benchmark in which

measurement uncertainty, institutional persistence, correlated country heterogeneity, and extreme growth observations are already addressed before any gain is attributed to nonlinearity (Table 8).

Table 8 M3R measurement, state, and linear growth parameters

Parameter	Mean	SD	5%	Median	95%
Loading: Voice and Accountability	0.803	0.010	0.786	0.803	0.820
Loading: Political Stability	0.729	0.011	0.711	0.729	0.747
Loading: Government Effectiveness	0.978	0.005	0.970	0.978	0.986
Loading: Regulatory Quality	0.942	0.006	0.932	0.942	0.952
Loading: Rule of Law	1.000	—	1.000	1.000	1.000
Loading: Control of Corruption	0.995	0.006	0.986	0.995	1.004
Institutional persistence	0.977	0.004	0.970	0.977	0.983
Institutional innovation SD	0.057	0.002	0.054	0.057	0.060
Linear governance slope	0.803	0.346	0.269	0.780	1.417
Lagged-growth coefficient	0.345	0.016	0.318	0.345	0.372
Student-t degrees of freedom	2.031	0.029	2.002	2.022	2.091

Notes: SD denotes posterior standard deviation. The 5% and 95% columns are the posterior quantiles and jointly form a 90% credible interval. Political Stability denotes Political Stability and Absence of Violence/Terrorism (PV), as

defined in Table 2. The Rule of Law is the loading anchor fixed at one; the institutional-persistence parameter is defined in Section 3.7. Table 8 reports the final audited M3R comparator posterior used in the locked same-baseline architecture, not an earlier development stage M3R fit.

6.4 Primary final audited M4R specification

The final audited M4R fit is the sole Primary Preferred Specification and the reference point for linked robustness comparisons. The model uses the 3,485-observation, 167-economy sample; all six WGI estimates and standard errors; the rule-of-law anchor; stationary latent governance dynamics; correlated country effects; centered year effects; lagged growth; five standardized lagged controls; a Student-t likelihood; and a smooth transition over the 5th-95th percentile admissible support. Development-stage Core M4R outputs were retained solely for provenance and were not incorporated into the final inferential narrative.

The preferred posterior supports regime differentiation but not a sharply located universal cutoff. The threshold median was 0.203, with a 95% credible interval of -1.295 to 1.152. The lower- and higher-governance slope medians are 1.121 and 0.511, respectively. Their higher-minus-lower contrast is -0.601 (95% credible interval: -1.206 to -0.116), and the posterior probability that the higher governance slope is smaller is 0.992. Therefore, H2 and H3 are supported by the slope contrast, whereas the broad transition-location posterior precludes the interpretation of 0.203 as a precise policy cutoff (Table 9).

Table 9 Primary audited M4R threshold results and computational diagnostics

Quantity	Posterior summary
Threshold	Median 0.203; 95% CrI -1.295 to 1.152
Lower-governance slope	Median 1.121; 90% CrI 0.545 to 1.855

Quantity	Posterior summary
Higher-governance slope	Median 0.511; 90% CrI -0.012 to 1.161
Higher-lower slope	Median -0.601; 95% CrI -1.206 to -0.116
P(higher slope < lower slope)	0.992
Transition speed	Median 7.752
Maximum structural R-hat	1.0026
Minimum structural ESS_bulk	1,698.3
Divergences / tree depth hits	0 / 0

Notes: CrI denotes credible interval; R-hat, rank-normalized potential scale-reduction diagnostic; ESS_bulk, bulk effective sample size. Threshold and higher-minus-lower rows report 95% CrIs; individual slopes report archived 5th and 95th quantiles (90% CrIs). All quantities were obtained from the locked final audited M4R posterior. The threshold is on the rule-of-law-anchored latent scale; lower and higher slopes are asymptotic coefficients on latent governance.

6.5 Audited nonlinear functional-form comparison

Functional form comparison qualifies the interpretation of M4R. Under the same measurement, state, country effect, control, and Student-t architecture, observation-level PSIS-LOO gives expected log predictive density (ELPD) values of -8,803.308 for M3Q, -8,804.796 for M4R, and -8,808.390 for M3R. M3Q exceeds M4R by 1.489 ELPD units (paired SE 0.605), while both nonlinear models outperform the latent linear benchmark in this comparison. Therefore, the evidence favors concavity over strict linearity but does not uniquely select a smooth-threshold functional form. Section 6.13 separately reports the final submission-level M4R-M3R same-baseline revalidation and country-grouped cross-validations (Table 10).

Table 10 Observation-level PSIS-LOO comparison under the audited architecture

Model	ELPD-LOO	SE	Difference from best	SE difference	Interpretation
M3Q latent quadratic	-8,803.308	77.055	0.000	0.000	Best, but close to M4R
M4R smooth threshold	-8,804.796	77.040	1.489	0.605	Near- equivalent to M3Q
M3R latent linear	-8,808.390	77.123	5.082	2.505	Lower predictive fit

Note: A smaller positive difference from the best model indicates a closer predictive performance. No Pareto-k value exceeded 0.7 in the three-model comparison. The small M3Q-M4R difference is interpreted as functional-form uncertainty rather than the decisive selection of one nonlinear model, consistent with the need to account for uncertainty in cross-validation comparisons (Bates et al., 2024; Sivula et al., 2025).

6.6 Parameter-recovery evidence

The final canonical parameter recovery experiment used 50 replications per scenario. For the higher-minus-lower slope contrast, the bias/root mean squared error (RMSE) was 0.053/0.686 under strong separation, 0.068/0.883 under weak separation, and 0.048/0.936 under the null; the 95% coverage was 1.00, 0.96, and 0.92, respectively. The threshold bias/RMSE was -0.001/0.216, 0.005/0.225, and 0.023/0.149, with 90% and 95% threshold coverage equal to 1.00 throughout. Under the null, the persistence bias/RMSE is -0.063/0.070 and 90%/95% coverage is 0.52/0.74. Therefore, the simulation supports the regime contrast more clearly than persistence, while reinforcing caution about the exact cutoff interpretation (Table 11).

Table 11 Final canonical parameter-recovery summary for the audited M4R architecture

Scenario	Target	True value	Bias	RMSE	90% coverage	95% coverage
Strong	Higher–lower slope	-1.100	0.053	0.686	0.94	1.00
Strong	Threshold	0.000	-0.001	0.216	1.00	1.00
Weak	Higher–lower slope	-0.400	0.068	0.883	0.94	0.96
Weak	Threshold	0.000	0.005	0.225	1.00	1.00
Null	Higher–lower slope	0.000	0.048	0.936	0.86	0.92
Null	Threshold	0.000	0.023	0.149	1.00	1.00
Null	Institutional persistence	0.850	-0.063	0.070	0.52	0.74

Note: Each scenario used 50 canonical replications. The recovery of the higher-minus-lower slope contrast is stronger than the recovery of latent-state persistence. The threshold interval coverage under the null should not be interpreted as substantive cutoff identification because the transition location is weakly meaningful when the true regime contrast is zero.

6.7 Temporal stability and crisis diagnostics

Temporal diagnostics do not support a single invariant calendar break in the conventional observed score relationship. The estimated governance slope change was positive for 2015-2023 (1.385; $p=0.007$), weaker for 2009-2023 and the 2020-2021 pandemic interval, and negligible for 2022-2023. Although the exploratory candidate-year scan reached its largest statistic in 2015, the Holm-adjusted p -value was 0.238. These results are best interpreted as evidence that conventional

governance-growth associations can vary over time rather than as alternative estimates of the latent M4R transition (Table 12).

Table 12 Fixed-date governance-slope stability diagnostics

Break definition	Slope before	Slope change	Slope after	p: governance change	p: joint slope change
2009–2023	4.020	1.038	5.058	0.098	0.0006
2015–2023	4.203	1.385	5.588	0.007	0.0003
2020–2021	4.506	1.559	6.064	0.071	0.104
2022–2023	4.647	0.171	4.818	0.748	0.318

Note: In each fixed-date row, “Slope change” is the estimated change in the conventional observed-score governance slope for the stated calendar window, and “Slope after” equals the corresponding pre-window slope plus that change. These are ancillary frequentist stability diagnostics, not alternative latent M4R threshold estimates.

6.8 Measurement-model sensitivity

Measurement model sensitivity indicates that negative-slope ordering is not generated by a single reliability scaling or a single factor anchor. Multiplying the reported WGI standard errors, changing the anchor to Government Effectiveness, and omitting individual WGI dimensions preserve a negative higher-minus-lower contrast. Because changing the anchor rescales the latent factor, cross-anchor comparisons focus on sign, ordering, credible separation, and boundary behavior rather than raw coefficient magnitudes. Table 13 therefore supports the robustness of qualitative regime ordering while preserving the scale dependence of the latent variable parameters.

Table 13 Audited measurement-model sensitivity

Specification	Threshold median	Lower slope	Higher slope	Higher–lower	P(higher<lower)
Primary audited M4R	0.203	1.121	0.511	-0.601	0.992
Government Effectiveness anchor	-0.010	1.244	0.572	-0.658	0.994
WGI SE × 0.75	0.207	0.813	0.268	-0.541	0.988
WGI SE × 1.25	0.139	1.299	0.659	-0.633	0.993
WGI SE × 1.50	-0.099	1.458	0.785	-0.677	0.994
Exclude Government Effectiveness	0.027	0.837	0.318	-0.506	0.981

Notes: The government effectiveness anchor model is not numerically scale-comparable with rule-of-law anchor models. The valid cross-anchor comparison concerns signs, ordering, credible separation, and boundary behavior. The WGI SE × 1.25 row reports the targeted final rerun (threshold 0.139; higher-minus-lower -0.633; P(higher<lower)=0.993), which passed the strict computational gate (maximum key R-hat 1.00143; minimum ESS_bulk 1,986.24; zero divergences, and zero maximum-treedepth hits).

6.9 Heterogeneous common shocks and lower-tail likelihood

Allowing richer common shocks strengthens, rather than removes, the negative asymptotic slope's ordering. The income-ranked-group-by-year specification yields a threshold median of 0.192,

lower- and higher-governance asymptotic slope medians of 1.207 and 0.544, and a higher-minus-lower median of -0.665 (95% credible interval: -1.298 to -0.153), with $P(\text{higher}<\text{lower})=0.995$. The observation-level PSIS-LOO favored the augmented common-shock model by 103.23 ELPD units (paired SE 15.26). The Step 16F two-piece Student-t specification also preserves the negative asymptotic slope ordering: the threshold median is 0.480 (95% credible interval: -1.175 to 1.172), the lower- and higher-governance asymptotic slope medians are 1.007 and 0.366, respectively, and the higher-minus-lower median is -0.636 (95% credible interval: -1.158 to -0.167), with $P(\text{higher}<\text{lower})=0.997$. Its observation-level ELPD-LOO was -8,797.321, an improvement of 7.475 ELPD units (paired SE 4.957) over the linked baseline M4R value of -8,804.796; no Pareto- k value exceeded 0.7. These checks reduce the concern that the primary contrast is driven solely by homogeneous year shocks or symmetric tail assumptions, but they do not create exogenous governance variation. The Step 16E and Step 16F predictive comparisons use their linked contemporaneous baseline M4R (ELPD = -8,804.796), rather than the Final Audited same-baseline M4R posterior reported in Table 18.

Table 14 Common-shock and lower-tail sensitivity

Model	Threshold	Lower slope	Higher slope	Higher–lower	P(higher<lower)	Predictive evidence
Primary audited M4R	0.203	1.121	0.511	-0.601	0.992	Reference
Income- ranked- group-by- year M4R	0.192	1.207	0.544	-0.665	0.995	Delta ELPD 103.23 (SE 15.26)

Model	Threshold	Lower slope	Higher slope	Higher–lower	P(higher<lower)	Predictive evidence
Two-piece	0.480	1.007	0.366	-0.636	0.997	Δ ELPD
Student-t						+7.475 (SE 4.957) vs linked M4R

Note. The Δ ELPD = 103.23 (paired SE = 15.26) entry for the income-ranked-group-by-year specification is retained from the linked Step 16E comparison against its contemporaneous baseline M4R (ELPD = -8,804.796). For the Step 16F two-piece Student-t specification, the threshold median was 0.480 (95% CrI -1.175 to 1.172), the higher-minus-lower median was -0.636 (95% CrI -1.158 to -0.167), and P(higher<lower)=0.997. Its ELPD-LOO was -8,797.321 versus -8,804.796 for the linked baseline, giving Δ ELPD = +7.475 (paired SE = 4.957), with no Pareto-k value above 0.7. Neither linked comparator is substituted for the Final Audited M4R primary specification; the Final Audited same-baseline M4R-M3R comparison is reported separately in Table 18 below.

6.10 Consecutive-coverage sensitivity

Changing the minimum consecutive history requirement from 6 to 12 years leaves the higher-minus-lower slope contrast negative across all specifications. The 8-year rule remains an audited baseline. However, the transition location is much less stable: under the 12-year rule, the median moves to -1.093, and 42.1% of the posterior threshold mass accumulates in the lowest 5% of admissible support. Therefore, the strict-coverage result reinforces slope ordering but weakens the identification of exact cutoffs because the transition becomes boundary-sensitive rather than newly precise (Table 15).

Table 15 Consecutive-coverage sensitivity

Minimum years	Observations	Economies	Threshold	Lower slope	Higher slope	Higher- lower	P(higher <lower)
6	3,498	169	0.315	1.103	0.473	-0.623	0.994
8 (primary)	3,485	167	0.203	1.121	0.511	-0.601	0.992
10	3,433	162	0.098	1.217	0.523	-0.691	0.997
12	3,402	159	-1.093	1.499	0.571	-0.914	0.998

Notes: The eight-year row reports the current Final audited M4R Primary Preferred Specification from Table 9. The six-, ten-, and twelve-year rows are separately re-estimated Step 16G coverage-rule sensitivity specifications. The historical Step 16G eight-year comparator reused an earlier preferred baseline and was retained only as provenance; it was not used as the current primary posterior in this table.

6.11 Governance timing sensitivity

Governance timing provides a more demanding test of whether contemporaneous contrast is purely alignment-specific. On an exact common sample of 3,149 observations from 167 economies, the higher-minus-lower median remains negative for contemporaneous, one-year-lagged, and two-year-lagged governance results. Posterior ordering probabilities fall from 0.988 to 0.969 and 0.930, and the 95% credible interval crosses zero in the lagged specifications. Therefore, the direction is stable, but the evidence becomes less decisive as the governance measure is moved farther back in time. Lagging should not be interpreted as resolving reverse causality; it changes temporal ordering without supplying the exogenous variation. Table 16 labels the three timing scenarios, T0–T2, to distinguish them from research hypotheses H1–H3.

Table 16 Lagged-governance timing sensitivity

Scenario	Timing	Threshold	Lower slope	Higher slope	Higher -lower	95% CrI	P(higher <lower)
T0	Contemporaneous	-0.031	1.099	0.510	-0.570	[-1.164, -0.069]	0.988
T1	Lag 1 year	-0.027	0.727	0.259	-0.469	[-1.046, 0.024]	0.969
T2	Lag 2 years	0.018	0.448	0.074	-0.370	[-0.915, 0.123]	0.930

6.12 Leave-one-WGI-dimension-out sensitivity

Leave-one-dimension-out re-estimation shows that no single WGI dimension is necessary for negative slope ordering. Among the models retaining the Rule of Law as the anchor, the contrast medians range from -0.506 to -0.631, and the ordering probabilities from 0.981 to 0.995. Omitting the Rule of Law requires re-anchoring on Government Effectiveness, so only the sign and ordering are comparable in that row. This result reduces concerns that the preferred nonlinear contrast is an artifact of a single governance component, while remaining consistent with a common-factor interpretation (Table 17).

Table 17 Leave-one-WGI-dimension-out sensitivity

Scenario	Omitted dimension	Threshold	Lower slope	Higher slope	Higher-lower	P(higher<lower)
I0	None	0.239	1.126	0.517	-0.608	0.993
I1	Voice and Accountability	0.325	1.144	0.509	-0.631	0.995
I2	Political Stability	0.377	1.063	0.458	-0.603	0.993

Scenario	Omitted dimension	Threshold	Lower slope	Higher slope	Higher–lower	P(higher<lower)
I3	Government Effectiveness	0.027	0.837	0.318	-0.506	0.981
I4	Regulatory Quality	0.031	0.963	0.378	-0.580	0.990
I5	Rule of Law; re-anchored	0.271	1.898	1.095	-0.807	0.999
I6	Control of Corruption	0.167	1.104	0.498	-0.599	0.994

Note: I0 is the separate all-six-dimension reference fit within Step 16I, not Table 9 primary posterior. I5 re-anchors on Government Effectiveness after omitting Rule of Law; therefore, its threshold and coefficient magnitudes are not directly scale-comparable. Political Stability denotes Political Stability and Absence of Violence/Terrorism (PV), as defined in Table 2.

6.13 Final same-baseline predictive, support, and prior-sensitivity evidence

Final submission-level revalidation under the locked audited architecture produced a deliberately mixed predictive conclusion. In the same-baseline M4R-M3R PSIS-LOO comparison, M4R had ELPD -8,804.978 and M3R -8,808.054, yielding a modest M4R-minus-M3R difference of 3.076 ELPD units (SE 2.1); neither model had a Pareto-k value above 0.7. Five-fold country-grouped validation reverses the aggregate sign: M4R totals -8,952.973 versus -8,948.834 for M3R, for a difference of -4.139, with a country-level SE of 6.953. M4R is favored in 95 of 167 economies, and the median country difference is 0.059; however, the aggregate grouped evidence does not establish a conditional transport advantage. Threshold-support re-estimation over 5%, 10%, and 15% trimmed supports preserves negative contrasts with medians from -0.602 to -0.581 and

P(higher<lower) from 0.991 to 0.993. All six source-audited prior scenarios also passed the strict computational gate, with contrast medians from -0.676 to -0.593 and ordering probabilities from 0.992 to 0.994. The near-hard transition shifts the threshold median to -1.003 while leaving the contrast negative, directly demonstrating that the slope ordering is more stable than the transition location (Tables 18 and 19).

Table 18 Final same-baseline predictive validation under the locked audited architecture

Validation	Final audited evidence	Interpretation
Observation-level PSIS-LOO: M4R vs M3R	M4R ELPD -8,804.978; M3R - 8,808.054; Δ ELPD (M4R–M3R) = +3.076; SE 2.1; Pareto-k >0.7 = 0 for both	Modest within-panel M4R advantage; not decisive
Five-fold country-grouped CV	M4R total ELPD -8,952.973; M3R -8,948.834; Δ = -4.139; country-level SE 6.953; M4R favored in 95/167 economies	No aggregate conditional country-level transport advantage
Functional-form context	M3Q exceeds M4R by 1.489 ELPD units (SE 0.605) in the linked three-model comparison	Concavity supported; threshold form not uniquely selected
Rolling-origin leave-future-out	No final audited output available	Not assessed; excluded from predictive-survival claim

Notes: Table 18 reports the final same baseline predictive revalidation. The grouped CV standard error is computed from economy-level M4R-M3R differences. Table 19 reports 95% posterior credible intervals. All three support fits, and all six prior sensitivity scenarios satisfy the strict computational gates; the prior scenarios have a maximum key R-hat no greater than 1.0062, a minimum bulk effective sample size (ESS)

of at least 832.1, zero divergences, and zero maximum tree depth hits. The near-hard-transition result is interpreted as a threshold-location sensitivity rather than a reversal of the regime-slope ordering.

Table 19 Final threshold-support and prior-sensitivity evidence under the locked audited M4R architecture

Scenario/check	Threshold median (95% CrI)	Higher–lower median (95% CrI)	P(higher<lower)
Threshold support: 5% trim	0.201 (-1.296, 1.168)	-0.602 (-1.189, -0.123)	0.993
Threshold support: 10% trim	0.403 (-1.000, 1.135)	-0.581 (-1.098, -0.104)	0.991
Threshold support: 15% trim	0.318 (-0.811, 1.016)	-0.595 (-1.116, -0.121)	0.993
Prior: slopes_tight	0.255 (-1.283, 1.175)	-0.593 (-1.140, -0.114)	0.993
Prior: slopes_diffuse	0.191 (-1.296, 1.157)	-0.615 (-1.214, -0.129)	0.993
Prior: measurement_tight	0.238 (-1.290, 1.162)	-0.602 (-1.190, -0.122)	0.993
Prior: measurement_diffuse	0.250 (-1.292, 1.153)	-0.602 (-1.205, -0.119)	0.992
Prior: persistence_diffuse	0.267 (-1.293, 1.167)	-0.611 (-1.180, -0.128)	0.994
Prior: near_hard_transition	-1.003 (-1.364, 1.148)	-0.676 (-1.355, -0.146)	0.994

The final locked-posterior PPC directly uses all 16,000 archived growth_rep draws from the four exact Final Audited M4R chains. The nominal 90% observation-level posterior predictive intervals covered 89.44% of the observed growth outcomes. The observed mean (1.888), median (2.171), SD (5.545), 99th percentile (13.649), sample extremes, and shares with $|\text{growth}| > 10, 20, \text{ and } 30$ all lie within the corresponding 95% posterior predictive distributions of replicated statistics. However, the observed 1st percentile (-15.088), 5th percentile (-5.962), and 95th percentile (8.333) fell outside the corresponding 95% replicated-statistic intervals. Year-level 90% coverage was also lower in 2020 (73.0%) and 2021 (69.3%). Therefore, the PPC supports overall calibration with a qualification for residual tail asymmetry and crisis-period misfit rather than an exact distributional fit.

6.14 Hypothesis and threshold-survival assessment

Taken together, the audited evidence supports H1, H2, and H3 but does not meet the full threshold for survival. H1 is supported because sampling and measurement propagation materially altered the cutoff location and classification uncertainty relative to the naïve 0.847 estimate. H2 is supported because the final M4R higher-minus-lower asymptotic slope contrast is credibly different from zero. H3 is strongly supported directionally because $P(\text{higher} < \text{lower}) = 0.992$, and the negative asymptotic slope ordering persists across the completed measurement, support, prior, common shock, likelihood, coverage, timing, and indicator composition checks. The strongest structural conclusion is nevertheless robust concavity and a smaller higher-governance asymptotic slope, not a unique threshold form or a demonstrated monotonic decline in the full point-wise marginal derivative. M3Q is predictively competitive, the transition location is diffuse and support-sensitive, and country-grouped validation provides no aggregate M4R advantage. The appropriate classification was weak structural threshold survival rather than full structural and predictive survival (Table 20).

Table 20 Integrated single-baseline inference and reporting decision

Inferential question	Final audited evidence	Decision
Does WGI uncertainty change the threshold inference?	Naïve cutoff 0.847; M2 medians 0.618 and 0.465 with broad intervals	H1 supported
Do the asymptotic regime slopes differ?	Primary higher–lower asymptotic slope median -0.601; 95% CrI -1.206 to -0.116	H2 supported

Inferential question	Final audited evidence	Decision
Is the higher governance asymptotic slope smaller?	P(higher asymptotic slope < lower asymptotic slope)=0.992	H3 strongly supported
Is the cutoff precisely determined?	Threshold 0.203; 95% CrI - 1.295 to 1.152; location shifts under support, near-hard transition, and 12-year coverage	No precise universal cutoff
Is the nonlinearity uniquely threshold-shaped?	M3Q exceeds M4R by 1.489 ELPD (SE 0.605) in the functional-form comparison	No; robust concavity, form uncertain
Does support/prior perturbation preserve the slope ordering?	Support-trim contrasts -0.602 to -0.581; six prior-sensitivity contrasts -0.676 to -0.593; all final strict gates pass	Yes; asymptotic slope ordering robust, threshold location sensitive
Does the structural evidence survive?	Negative ordering persists across exact linked measurement, support, prior, common-shock, likelihood, coverage, timing, and indicator-composition checks	Weak structural threshold survival
Does predictive evidence transport?	Same-baseline PSIS Δ ELPD +3.076 (SE 2.1), but grouped CV Δ ELPD -4.139 (country SE 6.953); M4R favored in 95/167 economies	No clear aggregate predictive survival
Is rolling-origin predictive survival established?	No final audited rolling-origin leave-future-out result	Not assessed

7. Discussion

7.1 Hypothesis overview and central interpretation

The three hypotheses are supported at different inferential levels. Measurement uncertainty changes the transition location and classification; the latent model retains a credible negative higher-minus-lower asymptotic slope contrast with $P(\text{higher} < \text{lower}) = 0.992$; and this ordering persists across completed robustness checks. However, the location remains diffuse, M3Q is slightly favored over M4R in the three-model observation-level comparison, and grouped validation gives M4R no aggregate advantage. The central M4R result is therefore a smaller higher-governance asymptotic slope, alongside broader evidence of concavity, not a precise cutoff or proof that the full pointwise marginal derivative declines monotonically across the observed support.

This distinction separates regime differentiation from cutoff precision. Bootstrap, measurement perturbation, support re-estimation, strict coverage, and the near-hard transition move the location while leaving slope ordering negative; the latent quadratic comparator adds functional-form uncertainty. Threshold research should therefore treat slope change, location precision, and functional-form uniqueness as separate propositions.

7.2 Measurement uncertainty and regime classification

H1 shows why measurement error matters especially when an estimated index defines regimes. The naïve cutoff of 0.847 shifts to M2 medians of 0.618 and 0.465 under alternative error-dependence assumptions, with broad intervals. Therefore, measurement uncertainty weakens the deterministic regime classification more clearly than it changes the negative slope ordering (Kaufmann & Kraay, 2024; Magnusson & Tarverdi, 2020; Qu et al., 2019).

The mechanism is nonlinear sorting with heterogeneous reliability: modest score revisions can reclassify observations near the transition, and unequal standard errors change their contribution to latent paths. This is consistent with measurement and decision theory, which favors propagating uncertainty into classification rather than suppressing it in point estimates (Ferrero & Petri, 2025; Levy & McNeish, 2025; Zamora & Towse, 2023).

7.3 Asymptotic slope ordering and competing institutional mechanisms

The joint support for H2 and H3, expressed as a smaller higher-governance asymptotic slope, is more consistent with saturation than with increasing complementarity in the annual growth. Institutional improvements may relax binding enforcement, administrative, regulatory, and corruption constraints at lower governance levels, whereas comparable improvements can remain socially valuable but be associated with smaller growth changes in the higher governance asymptote once basic bottlenecks are less binding. This is a mechanism-consistent interpretation of the estimated asymptotic slope ordering, not direct evidence that the pointwise marginal effect declines at every value of latent governance within a smooth transition.

This interpretation qualifies rather than overturns prior evidence on nonlinear governance. Existing studies report regime dependence and heterogeneity across dimensions and development groups (Asamoah, 2021; Glawe, 2025; Misi Lopes et al., 2023; Ochi et al., 2023, 2024; Parsa & Datta, 2023). The present results add a measurement boundary: slope differentiation can survive uncertainty propagation even when the transition score is imprecise and a quadratic form remains competitive.

Economic significance should be interpreted on the rule-of-law-anchored latent scale. Re-anchoring changes raw coefficient magnitudes even when slope ordering is preserved, and the manuscript does not report audited percentile-shift contrasts. Policy interpretation should therefore

emphasize the relative lower-versus-higher slope ordering rather than mechanically converting latent units into universal policy increments.

7.4 Predictive transportability and policy implications

Predictive evidence limits structural claims. Same-baseline PSIS-LOO gives M4R a modest 3.076 ELPD advantage over M3R (SE 2.1), whereas country-grouped validation yields -4.139 (SE 6.953); M3Q exceeds M4R by 1.489 ELPD (SE 0.605). The final locked-posterior PPC is an in-sample adequacy diagnostic rather than transport evidence: its nominal 90% observation-level predictive intervals cover 89.44% of the observed outcomes, but residual tail-quantile and crisis-period misfit remain. These design-specific results support within-sample concavity more clearly than the predictive superiority of a threshold rule across economies (Bates et al., 2024; Sivula et al., 2025).

Policy use should match this uncertainty. The posterior median 0.203 should not be treated as a minimum WGI target or universal reform trigger. More defensible reporting combines transition intervals, regime probabilities, measurement/prior sensitivity, and the predictive validation target, while country-specific reform priorities remain tied to diagnosed institutional bottlenecks rather than mechanical distance from a global cutoff.

7.5 Theoretical, methodological, empirical, and policy contributions

The theoretical contribution is the distinction between surviving asymptotic slope ordering and surviving cutoff location. A saturation mechanism can be consistent with a smaller higher-governance asymptotic slope, even when the transition is diffuse, support-sensitive, and not predictively transportable. Therefore, institutional nonlinearity should be theorized in terms of changes in the governance growth slope structure rather than necessarily as a sharp tipping point;

within the M4R transition region, the pointwise derivative of the level contribution additionally depends on the derivative of the logistic transition weight.

Methodologically, the study integrates six WGI indicators, producer-reported standard errors, a persistent latent state, correlated country effects, a robust growth likelihood, and smooth regime differentiation within one Bayesian architecture. Matched latent linear and quadratic comparators separate measurement correction from functional form, while targeted robustness checks distinguish sampling, measurement, residual specification, posterior dependence, and predictive transportability. The contribution is an auditable applied architecture rather than a new estimator.

Empirically, the provenance-controlled result is stable slope ordering but fragile location. Across support and prior scenarios, the higher-minus-lower contrast remains around -0.6 with ordering probability above 0.99; measurement, common-shock, tail, coverage, timing, and dimension checks do not reverse its sign. In contrast, location shifts under strict coverage and near-hard transition, and grouped prediction does not favor M4R in aggregate.

The policy contribution is therefore uncertainty-aware threshold reporting. Studies using estimated institutional indices should disclose measurement assumptions, transition uncertainty, boundary behavior, classification ambiguity, functional-form alternatives, and predictive targets rather than treating estimated threshold crossing as deterministic.

7.6 Limitations and future research

Several limitations remain. The common-factor model may obscure dimension-specific structures; anchor choice affects latent scale; common support can induce selection; lagged specifications are less decisive; recovery is weaker for persistence; and 27 of 150 simulations required initialization-only stabilization. M3Q slightly outperforms M4R in the three-model PSIS comparison, country-

grouped validation gives no aggregate M4R advantage, and rolling origin validation is unavailable. The final locked-posterior PPC is close to the nominal overall coverage but shows residual tail-quantile misfit and weaker calibration in 2020-2021 and in some economies. The observational design also does not eliminate reverse causality or time-varying confounding, and the asymptotic slope contrast does not establish a monotonic decline in the full pointwise marginal derivative across the transition region.

Future work should prioritize inferential limitations rather than additional specification counts. Longer panels can separate short-run adjustment from persistent change; multi-factor or dimension-specific models can test the common-factor assumption; correlated measurement errors and alternative institutional datasets can strengthen measurement validation; rolling-origin and fully unconditional new-country prediction can assess transportability; and stronger causal claims require external instruments, quasi-experimental reforms, discontinuities, or other exogenous institutional shocks.

8. Conclusion

After propagating measurement uncertainty through the locked final audited M4R, the higher-minus-lower asymptotic governance slope remains credibly negative across completed measurement, support, prior, common-shock, likelihood, coverage, timing, and indicator composition checks. The transition location is much less stable, with a broad posterior and sensitivity to support, transition sharpness, and strict coverage. Therefore, the evidence supports a credibly smaller higher-governance asymptotic slope and broader concavity more clearly than a precise institutional cutoff; it does not establish that the full pointwise marginal derivative declines monotonically throughout the smooth transition.

Functional-form and predictive evidence further limit the threshold claim. M3Q is slightly favored over M4R in the three-model PSIS-LOO comparison; final same-baseline PSIS-LOO gives M4R only a modest advantage over M3R; and country-grouped validation gives no aggregate M4R advantage. The appropriate classification is weak structural threshold survival, not a universal cutoff, portable prediction rule, or causal governance effect.

The broader contribution is a shift from deterministic threshold crossing to probabilistic regime inference. When an estimated index both enters the outcome equation and defines regimes, measurement uncertainty is part of the estimand. Governance applications should therefore report transition regions, regime probabilities, support/prior sensitivity, and model-comparison uncertainty alongside cutoff estimates; the same principle applies to other latent or composite indices used for policy classification.

The conclusions remain conditional on the observational design, common factor representation, WGI error model, sample selection rules, functional form, incomplete temporal validation, and residual tail/crisis-period misfit detected by the final locked-posterior PPC. Future research should add stronger external institutional variation, richer measurement structures, time-varying effects, fully unconditional predictions, and direct posterior summaries of pointwise marginal derivatives, where substantive marginal-effect claims are required. Reliable institutional policy requires quantifying not only whether governance growth slopes differ, but also how the apparent threshold and any pointwise marginal effect depend on measurement, support, functional form, and predictive task.

Author Contributions

The author was responsible for the conceptualization, methodology, data curation, formal analysis, software, validation, visualization, writing of the original draft, writing of the review and editing, and preparation of the replication materials.

Competing interests

The authors declare that there are no known competing financial interests or personal relationships that could have influenced the work reported in this study.

Declaration of AI Use

The authors used artificial intelligence (AI) tools only for language editing, academic style refinement, formatting assistance and consistency checking. The AI tool was not used to generate data, perform statistical analyses, fabricate results, create references, or make substantive theoretical or empirical decisions. The authors reviewed and verified all scientific claims, model specifications, numerical results, citations, interpretations, and final manuscript content and took full responsibility for the manuscript.

Ethical Approval

Ethical approval was not required for this study because it did not involve human or animal participants. This research relied exclusively on country-level macroeconomic and institutional panel data sourced from publicly accessible databases.

Informed Consent

Informed consent was not required because the study did not involve human subjects. All analyses were conducted using aggregated secondary data from the national level.

References

- Acemoglu, D., Johnson, S., & Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review*, 91(5), 1369–1401. <https://doi.org/10.1257/aer.91.5.1369>
- Adin, A., Krainski, E. T., Lenzi, A., Liu, Z., Martínez-Minaya, J., & Rue, H. (2024). Automatic cross-validation in structured models: Is it time to leave out leave-one-out? *Spatial Statistics*, 62, Article 100843. <https://doi.org/10.1016/j.spasta.2024.100843>
- Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*, 58(2), 277–297. <https://doi.org/10.2307/2297968>
- Asamoah, L. A. (2021). Institutional quality and income inequality in developing countries: A dynamic panel threshold analysis. *Progress in Development Studies*, 21(2), 123–143. <https://doi.org/10.1177/14649934211016715>
- Baltagi, B. H., Bresson, G., Chaturvedi, A., & Lacroix, G. (2022). Robust dynamic panel data models using ε -contamination. In A. Chudik, C. Hsiao, & A. Timmermann (Eds.), *Essays in honor of M. Hashem Pesaran: Panel modeling, micro applications, and econometric methodology* (Advances in Econometrics, Vol. 43B, pp. 307–336). Emerald Publishing Limited. <https://doi.org/10.1108/S0731-90532021000043B013>
- Bates, S., Hastie, T., & Tibshirani, R. (2024). Cross-validation: What does it estimate and how well does it do it? *Journal of the American Statistical Association*, 119(546), 1434–1445. <https://doi.org/10.1080/01621459.2023.2197686>

- Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1), 115–143. [https://doi.org/10.1016/S0304-4076\(98\)00009-8](https://doi.org/10.1016/S0304-4076(98)00009-8)
- Cooper, A., Vehtari, A., Forbes, C., Simpson, D., & Kennedy, L. (2024). Bayesian cross-validation by parallel Markov chain Monte Carlo. *Statistics and Computing*, 34, Article 119. <https://doi.org/10.1007/s11222-024-10404-w>
- Ferrero, A., & Petri, D. (2025). Measurement uncertainty and information enabled decision making. *IEEE Open Journal of Instrumentation and Measurement*, 4, 1–9. <https://doi.org/10.1109/OJIM.2025.3626105>
- Glawe, L. (2025). Nonlinearities in the institutions-growth relationship in a dynamic panel data framework: Evidence from China's provinces. *China Economic Review*, 94, Article 102583. <https://doi.org/10.1016/j.chieco.2025.102583>
- Griliches, Z., & Hausman, J. A. (1986). Errors in variables in panel data. *Journal of Econometrics*, 31(1), 93–118. [https://doi.org/10.1016/0304-4076\(86\)90058-8](https://doi.org/10.1016/0304-4076(86)90058-8)
- Hansen, B. E. (1999). Threshold effects in non-dynamic panels: Estimation, testing, and inference. *Journal of Econometrics*, 93(2), 345–368. [https://doi.org/10.1016/S0304-4076\(99\)00025-1](https://doi.org/10.1016/S0304-4076(99)00025-1)
- Hirano, K. (2002). Semiparametric Bayesian inference in autoregressive panel data models. *Econometrica*, 70(2), 781–799. <https://doi.org/10.1111/1468-0262.00305>
- Karagiannis, G., & Paleologou, S.-M. (2025). Governance and economic growth in Africa: Evidence from linear, nonlinear and dynamic panel analysis. *Empirical Economics*, 69(1), 39–75. <https://doi.org/10.1007/s00181-025-02726-z>

- Kaufmann, D., & Kraay, A. (2024). *The Worldwide Governance Indicators: Methodology and 2024 update* (Policy Research Working Paper No. 10952). World Bank.
<https://doi.org/10.1596/1813-9450-10952>
- Kremer, S., Bick, A., & Nautz, D. (2013). Inflation and growth: New evidence from a dynamic panel threshold analysis. *Empirical Economics*, 44(2), 861–878.
<https://doi.org/10.1007/s00181-012-0553-9>
- Kruschke, J. K. (2021). Bayesian analysis reporting guidelines. *Nature Human Behaviour*, 5, 1282–1291. <https://doi.org/10.1038/s41562-021-01177-7>
- Levy, R., & McNeish, D. (2025). Measurement and uncertainty preserving parametric modeling for continuous latent variables with discrete indicators and external variables. *Journal of Educational and Behavioral Statistics*, 50(2), 239–271.
<https://doi.org/10.3102/10769986241254348>
- Magnusson, L. M., & Tarverdi, Y. (2020). Measuring governance: Why do errors matter? *World Development*, 136, Article 105061. <https://doi.org/10.1016/j.worlddev.2020.105061>
- Misi Lopes, L. E., Packham, N., & Walther, U. (2023). The effect of governance quality on future economic growth: An analysis and comparison of emerging market and developed economies. *SN Business & Economics*, 3(6), Article 108. <https://doi.org/10.1007/s43546-023-00488-3>
- Mundlak, Y. (1978). On the pooling of time series and cross-section data. *Econometrica*, 46(1), 69–85. <https://doi.org/10.2307/1913646>
- North, D. C. (1990). *Institutions, institutional change and economic performance*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511808678>

- Ochi, A., Saidi, Y., & Labidi, M. A. (2023). Non-linear threshold effect of governance quality on economic growth in African countries: Evidence from panel smooth transition regression approach. *Journal of the Knowledge Economy*, 14(4), 4707–4729.
<https://doi.org/10.1007/s13132-022-01084-w>
- Ochi, A., Saidi, Y., & Labidi, M. A. (2024). Nonlinear threshold effect of governance quality on poverty reduction in South Asia and Sub-Saharan Africa: A dynamic panel threshold specification. *Journal of the Knowledge Economy*, 15(1), 4239–4264.
<https://doi.org/10.1007/s13132-023-01271-3>
- Ogbuabor, J. E., Emeka, E. T., & Ogbuabor, C. A. (2025). Effects of entrepreneurship and governance quality on global and regional economic performance: A pathway to sustainable development. *Sustainable Development*, 33(2), 2842–2863.
<https://doi.org/10.1002/sd.3267>
- Parsa, M., & Datta, S. (2023). Institutional quality and economic growth: A dynamic panel data analysis of MICs and HICs for 2000–2020. *International Economic Journal*, 37(4), 675–712. <https://doi.org/10.1080/10168737.2023.2261012>
- Qu, G., Slagter, B., Sylwester, K., & Doiron, K. (2019). Explaining the standard errors of corruption perception indices. *Journal of Comparative Economics*, 47(4), 907–920.
<https://doi.org/10.1016/j.jce.2019.07.003>
- Razzaq, A., Sharif, A., Öztürk, I., & Afshan, S. (2023). Dynamic and threshold effects of energy transition and environmental governance on green growth in COP26 framework. *Renewable and Sustainable Energy Reviews*, 179, Article 113296.
<https://doi.org/10.1016/j.rser.2023.113296>

- Seo, M. H., & Shin, Y. (2016). Dynamic panels with threshold effect and endogeneity. *Journal of Econometrics*, 195(2), 169–186. <https://doi.org/10.1016/j.jeconom.2016.03.005>
- Sinha, S. (2021). Bayesian approaches for handling covariate measurement error. In G. Y. Yi, A. Delaigle, & P. Gustafson (Eds.), *Handbook of measurement error models* (pp. 527–554). Chapman & Hall/CRC. <https://doi.org/10.1201/9781315101279-24>
- Sivula, T., Magnusson, M., Alonzo Matamoros, A., & Vehtari, A. (2025). Uncertainty in Bayesian leave-one-out cross-validation based model comparison. *Bayesian Analysis*. Advance online publication. <https://doi.org/10.1214/25-BA1569>
- Skarstein, E. S., Martino, S., & Muff, S. (2023). A joint Bayesian framework for missing data and measurement error using integrated nested Laplace approximations. *Biometrical Journal*, 65(8), Article e2300078. <https://doi.org/10.1002/bimj.202300078>
- Winter, S. D., & Depaoli, S. (2023). Illustrating the value of prior predictive checking for Bayesian structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 30(6), 1000–1021. <https://doi.org/10.1080/10705511.2022.2164286>
- Yang, L., Chen, I.-P., Lee, C., & Ren, M. (2024). Panel threshold model with covariate-dependent thresholds and unobserved individual-specific threshold effects. *Econometric Reviews*, 43(7), 452–489. <https://doi.org/10.1080/07474938.2024.2339147>
- Zamora, B., & Towse, A. (2023). The cost-per-QALY threshold in England: Identifying structural uncertainty in the estimates. *Frontiers in Health Services*, 2, Article 936774. <https://doi.org/10.3389/frhs.2022.936774>

Zhao, X., Lu, S., & Yuan, S. (2023). How does the digitization of government environmental governance affect environmental pollution? Spatial and threshold effects. *Journal of Cleaner Production*, 415, Article 137670. <https://doi.org/10.1016/j.jclepro.2023.137670>

Appendix A. Reproducible Estimation and Validation Protocol

The replication protocol begins with a 3,485-observation, model-ready panel and locks in the final audited M4R configuration before any robustness comparisons. M0 and M1 are estimated on the common baseline sample, followed by 2,000 country bootstrap replications and 2,000-draw composite measurement-error exercises. The final audited M3R, M3Q, and M4R fits used the same measurement, state, country effect, control, and Student-t architecture. The primary M4R posterior was identified by the rule-of-law intercept/loading anchor, 5th–95th support, four chains, and strict convergence gates. The executable specification is source-audited: free WGI intercepts use $\text{Normal}(0, 1)$, free loadings $\text{Normal}(1, 0.5)$, the Rule of Law is fixed at intercept 0 and loading 1, persistence is mapped through an inverse-logit transformation, and transition speed is positive with a $\text{Lognormal}(\log 8, 0.35)$ prior.

Parameter recovery uses 50 replications for each of the strong, weak, and null regime separations. The final canonical set retained 123 primary fits, 22 diagnostic-triggered anchor-tau rescues, and 5 further prior-centered latent-state rescues; rescue changes initialization only, reuses the exact simulated data, and leaves the locked model unchanged, while historical recovery outputs remain provenance only. Exact linked robustness checks alter one assumption at a time: measurement anchor or WGI uncertainty scale, income-ranked-group-by-year shocks, lower-tail likelihood, consecutive coverage, governance timing, and indicator inclusion. After the posterior lock, PSIS-LOO, five-fold country-grouped cross-validation, threshold-support sensitivity, and six prior-sensitivity scenarios were finalized under the audited architecture; the two measurement-prior

scenarios were source-repaired and rerun without overwriting successful provenance. The final submission-level PPC was then finalized by directly extracting all 16,000 archived growth_rep generated quantities from the four exact locked Final Audited M4R chains, without Stan recompilation, model refitting, or replacement predictive simulation. The Development-stage Core M4R and reconstructed PPC artifacts remain provenance only. Rolling-origin leave-future-out validation is not available. Every reported numerical result is traced to the final audited posterior or an explicitly linked re-estimation or diagnostic artifact.

Appendix B. Final Audited Computational and Validation Evidence

B.1 Primary audited baseline

The Primary Preferred Specification is the Rule-of-Law-anchored M4R fit used in the final M3R–M3Q–M4R comparison and as the comparator for the exactly linked robustness analyses. It contains 3,485 observations from 167 economies and passes the strict gate with a maximum key R -hat of 1.0026, a minimum key bulk ESS of 1,698.3, zero divergences and zero tree depth hits.

B.2 Posterior quantities

The audited threshold median was 0.203 (95% credible interval: -1.295 to 1.152). The lower- and higher-governance slope medians are 1.121 and 0.511, respectively. The higher-minus-lower median was -0.601 (95% credible interval: -1.206 to -0.116), and $P(\text{higher} < \text{lower}) = 0.992$.

B.3 Functional-form validation

The three-model observation-level PSIS-LOO functional-form comparison is -8,803.308 for M3Q, -8,804.796 for M4R, and -8,808.390 for M3R; M3Q exceeds M4R by 1.489 ELPD units (paired SE 0.605), supporting concavity without uniquely selecting the threshold. In the final same-baseline M4R–M3R PSIS revalidation, M4R had ELPD -8,804.978 versus -8,808.054 for M3R, a

modest 3.076-unit advantage (SE 2.1), with no Pareto-k above 0.7. Five-fold country-grouped validation yields M4R -8,952.973 and M3R -8,948.834, so M4R-minus-M3R = -4.139 (country-level SE 6.953); 95 of 167 economies favor M4R, but there is no aggregate conditional transport advantage.

B.4 Parameter recovery

Across 50 canonical replications per scenario, the bias/RMSE/95% coverage for the higher-minus-lower slope contrast was 0.053/0.686/1.00, 0.068/0.883/0.96, and 0.048/0.936/0.92 under strong, weak, and null separations, respectively. The threshold bias/RMSE were -0.001/0.216, 0.005/0.225, and 0.023/0.149, with 90% and 95% coverage equal to 1.00 throughout. Under the null, persistence bias/RMSE is -0.063/0.070 and 90%/95% coverage is 0.52/0.74; null-threshold coverage is not interpreted as a cutoff identification.

B.5 Replication boundary

Development-stage Core M4R summaries and incompatible legacy predictive, support, prior-sensitivity, posterior-predictive, and parameter-recovery outputs remain in the versioned archive; historical recovery results remain provenance only, and Table 11 is generated exclusively from the clean canonical 150-dataset. The final same-baseline PSIS-LOO, country-grouped CV, threshold support, six-scenario prior-sensitivity evidence, and direct locked-posterior PPC were used as submission evidence. The original failed measurement-prior runs remain provenance only; repaired Step 16N measurement_tight and Step 16N-R V5 measurement_diffuse replace them for the final inference. Development-stage and reconstructed PPC artifacts remain provenance only; the authoritative PPC uses the original archived growth_rep-generated quantities from all 16,000

retained draws in the four exact locked Final Audited M4R chains. Rolling-origin leave-future-out validation remains unassessed and is not used to support predictive survival claims.

B.6 Final locked-posterior predictive check

The submission-level PPC uses the original archived growth_ rep-generated quantities from all 16,000 retained draws in the four exact locked Final Audited M4R chains. The nominal 90% observation-level predictive intervals covered 89.44% of the observed outcomes. The observed mean, median, standard deviation, 99th percentile, sample extremes, and the shares of observations with absolute growth above 10, 20, and 30 fell within the corresponding 95% distributions of replicated statistics. The observed 1st percentile (-15.088) was below the replicated 95% range (-13.488 to -9.954), the observed 5th percentile (-5.962) was below its range (-5.781 to -4.841), and the observed 95th percentile (8.333) was slightly below its range (8.351 to 9.184). Coverage was lower in 2020 (73.0%) and 2021 (69.3%). The final PPC is therefore interpreted as an overall calibration with a residual tail and crisis-period misfit, not as evidence of an exact distributional fit or cross-economy transportability.

Appendix C. Completed Additional Q1 Validation Results

C.1 Measurement-model sensitivity

Alternative anchors, WGI standard error multipliers, and individual dimension exclusions preserve the negative higher-minus-lower slope ordering. Anchor changes are interpreted by sign and ordering, not by raw magnitude.

C.2 Heterogeneous common shocks and lower-tail likelihood

The income-ranked-group-by-year model yielded a contrast of -0.665 (95% credible interval: -1.298 to -0.153), with $P(\text{higher} < \text{lower}) = 0.995$. The two-piece Student-t likelihood preserves the ordering and modestly improves the lower-tail prediction.

C.3 Consecutive coverage

Using the current Final audited M4R primary posterior as the eight-year reference and the separately re-estimated Step 16G six-, ten-, and twelve-year coverage specifications, the higher-minus-lower contrast remains negative, ranging from -0.601 to -0.914. The twelve-year rule places 42.1% of the threshold mass in the lowest 5% of the support.

C.4 Governance timing

On the exact 3,149-observation common sample, the contrast medians are -0.570, -0.469, and -0.370 for contemporaneous, one-year, and two-year timing. Posterior certainty declines as the lag length increases.

C.5 Leave-one-dimension-out

Every WGI omission preserves a negative order. Excluding Government Effectiveness attenuates the contrast the most; excluding Rule of Law requires re-anchoring and is not numerically comparable across the scales.

C.6 Integrated interpretation

The final evidence supports robust concavity, stable negative asymptotic slope ordering, and weak cutoff identification. Threshold-support medians shift across the 5%, 10%, and 15% admissible supports, but the higher-minus-lower asymptotic slope contrast remains approximately -0.60, with $P(\text{higher} < \text{lower})$ above 0.99. All six source-audited prior scenarios pass strict computational gates; their contrast medians range from -0.676 to -0.593, with $P(\text{higher} < \text{lower})$ between 0.992 and 0.994, whereas the near-hard transition shifts the threshold median to -1.003. The direct locked-posterior PPC uses all 16,000 archived growth_rep draws and gives 89.44% coverage for nominal 90% observation-level predictive intervals; global moments, extremes, and extreme-event shares

are broadly reproduced, but the observed 1st, 5th, and 95th percentiles reveal residual distributional misfit, and calibration is weaker in 2020-2021. This PPC evidence concerns in-sample generative adequacy, not out-of-sample transportability. The same-baseline PSIS-LOO gives M4R a modest within-panel advantage over M3R, but five-fold country-grouped validation provides no aggregate predictive advantage. Rolling-origin leave-future-out validation has not yet been assessed. Therefore, the integrated classification remains a weak structural threshold survival model rather than a fully predictive survival model.