# How Robust Are the Legatum Prosperity Index Rankings?

## A Monte Carlo and Sensitivity Analysis of Weighting, Normalization, and Aggregation

**Adrian Erlikhman, Ryan Erlikhman** Los Angeles Center for Enriched Studies, Los Angeles, CA

---

## Abstract

The Legatum Prosperity Index ranks 167 countries by an equal-weighted mean of twelve pillar scores. We audit how much of the 2023 ranking survives when this construction is treated as one point in a space of defensible alternatives. We first verify that the equal-weighted mean of the published pillar scores reproduces every official score to within $5 \times 10^{-11}$ and every official rank without exception, so all downstream rank movement is attributable to our perturbations alone. We then propagate weight uncertainty through 10,000 Monte Carlo draws under three priors of increasing agnosticism, and run a global variance-based sensitivity analysis (Sobol' indices, Saltelli design, 16,384 evaluations) over fourteen inputs: the twelve pillar weights plus categorical selectors for normalization and aggregation. The ranking is globally stable, with median Spearman correlation to the published order never below 0.987 under any prior, but locally soft: the median country's 90 percent rank interval spans 22 places under agnostic weights, widening to a median of 37 places for ranks 68–100 and a maximum of 55 (Turkmenistan). The Sobol' decomposition concentrates rank variance in the weights on Personal Freedom (S_T = 0.23) and Natural Environment (S_T = 0.15), and across pillars a weight's total effect is strongly anticorrelated with its pillar's main-effect $\eta^2$ against the composite (Spearman $-0.76$, $p = 0.004$): the weights that matter most belong to the pillars least correlated with the aggregate. The aggregation selector's total effect exceeds nine of the twelve individual weights, and the normalization selector draws 65 percent of its total effect from interactions. Equal weighting is therefore a substantive decision about how much independent information the environmental and civil-liberty pillars contribute, not a neutral default.

**Keywords:** composite indicators, Legatum Prosperity Index, Monte Carlo simulation, Sobol' indices, sensitivity analysis, rank robustness, country rankings

---

## 1. Introduction

Composite indices reduce hundreds of indicators into one number per country. That number travels into policy documents, league-table journalism, and cross-country regressions that treat the index as data. Every such index embeds three construction choices its users rarely see: how indicators are normalized, how components are weighted, and how the weighted components are aggregated. Saisana et al. (2005) showed that these choices can move country ranks by enough to reverse the conclusions drawn from an index, and the OECD–JRC handbook (2008) has since made

uncertainty and sensitivity analysis a recommended step of composite-indicator construction. Producers rarely publish one.

We provide such an analysis for the 2023 edition of the Legatum Prosperity Index. The Legatum index is a good audit target on three grounds. It is widely cited, covering 167 countries annually since 2007. Its top-level construction is exactly documented, an equal-weighted arithmetic mean of twelve pillar scores (Legatum Institute, 2023), which permits exact replication before any perturbation, a precondition most composite audits cannot meet. And equal weighting poses a sharp question: assigning every pillar 1/12 looks neutral, but nominal weights and effective importance are known to diverge in composites (Paruolo et al., 2013; Becker et al., 2017), and whether the divergence is material here is an empirical question with a quantitative answer.

This paper contributes four things. First, we verify exact replication: the equal-weighted mean of the published pillar scores reproduces all 167 official ranks with score deviation below $5 \times 10^{-11}$, ruling out undocumented postprocessing at the pillar-to-index step. Second, we propagate 10,000 weight draws under three priors of increasing agnosticism (near-equal, a policy-adjustment grid, and fully uniform) and report per-country 90 percent rank intervals, finding a median width of 22 places under the uniform prior with a strongly tier-dependent profile. Third, we run a 14-input Sobol' analysis attributing rank variance jointly to weights, normalization, and aggregation; two weights dominate, and the aggregation rule out-ranks nine of twelve individual weights. Fourth, we quantify a divergence between assigned weights and realized importance: the Sobol' total effect of a pillar's weight is strongly anticorrelated with that pillar's realized importance $\eta^2$ against the composite. The weights that most move the ranking protect precisely the information the consensus of the other pillars does not carry.

---

## 2. Related Work

**Uncertainty and sensitivity analysis of composites.** The methodological template we follow is due to Saisana et al. (2005): treat normalization, weighting, and aggregation as uncertain modeling assumptions, propagate them by Monte Carlo, and decompose the variance of output ranks with global sensitivity indices. The OECD–JRC handbook (2008) codified this workflow and catalogued the standard alternatives per construction step. Applications to live, widely cited country rankings remain scarce, and producers do not release uncertainty statements with their league tables.

**Variance-based sensitivity analysis.** The functional-ANOVA framework originates with Sobol' (2001). Saltelli et al. (2010) supply the sampling designs and estimators that make first-order and total-effect indices computable at $N(k + 2)$ model evaluations, and Herman and Usher (2017) implement them in SALib, which we use. Our model output is a rank, a discrete function of the inputs; the decomposition applies without modification since it requires only square-integrability.

**Weights versus importance.** Paruolo et al. (2013) established that the nominal weight of a component is a poor proxy for its effective influence, and proposed the main-effect correlation ratio $\eta^2$ as the appropriate importance measure. Becker et

al. (2017) give methods for closing the gap between assigned weights and realized importance. We adopt $\eta^2$ and contribute the complementary observation, quantified for this index, that a weight's leverage over the ranking is itself governed by its pillar's $\eta^2$. Leverage concentrates where correlation with the composite is weakest.

---

## 3. Data and Baseline Replication

The 2023 Legatum Prosperity Index scores 167 countries through a three-level architecture. Three hundred indicators are normalized by a distance-to-frontier transformation and combined into 67 elements. Elements aggregate into twelve pillars, each on a 0–100 scale: Safety and Security, Personal Freedom, Governance, Social Capital, Investment Environment, Enterprise Conditions, Infrastructure and Market Access, Economic Quality, Living Conditions, Health, Education, and Natural Environment. Pillars then average with equal weights into the overall score (Legatum Institute, 2023). We take the published table of pillar scores as our data (167 countries × 12 pillars) and perturb only the pillar-to-index step. Everything below the pillar level is held at published values, a scoping decision we return to in Section 7.

Under a weight vector w = (w_1, ..., w_{12}) whose entries are non-negative and sum to 1, country i's composite score is the weighted sum of its twelve pillar scores, and its rank is one plus the number of countries with a higher composite score. The published index is the special case where every pillar receives weight 1/12.

Computed from the public data file, this baseline matches the official overall score for every country with maximum absolute deviation $5 \times 10^{-11}$, and reproduces every official rank (Spearman and Kendall correlations of exactly 1.0 over 167 positions). Replication is exact rather than approximate. An audit that cannot replicate its baseline confounds reconstruction error with genuine sensitivity, and every rank interval reported below is free of that confound.

---

## 4. Methods

### 4.1 Weight-uncertainty propagation

The first exercise varies only the weights, holding normalization at the native scale and aggregation additive. This isolates the question most readers ask of an equal-weighted index: do the weights matter? Any perturbation scheme encodes a position on how wrong equal weights might plausibly be, so we bracket the space of positions with three priors and draw M = 10,000 weight vectors from each.

The **uniform prior** draws weight vectors at random from every possible set of twelve weights that sums to 1, letting any single pillar dominate or vanish. It is fully agnostic and serves as a stress test. Formally this is the flat Dirichlet distribution (Appendix A).

The **near-equal prior** keeps every weight close to 1/12, with a standard deviation of about 0.011. It represents an analyst who accepts equal weighting but concedes it

cannot be exactly right. Formally this is a concentrated Dirichlet.

The **grid prior** mimics how index designers actually adjust weights when they depart from equal: it draws each raw weight independently from the set {0.5, 1, 1.5, 2}, then rescales all twelve so they sum to 1. We consider this the most policy-relevant of the three priors.

### 4.2 Joint perturbation design

The second exercise opens all three construction choices at once. The Sobol' sampling design used below requires inputs on a unit hypercube, so we work with fourteen coordinates in [0, 1]: twelve that generate weights and two that select a normalization and an aggregator. The twelve weight-coordinates are transformed into a weight vector on the simplex by a standard map that reproduces the uniform prior from Section 4.1 exactly (Appendix A). The remaining two coordinates each pick one option from a discrete menu:

- **Normalizations** (four options): the native distance-to-frontier scale, min–max rescaling to [0, 100], z-scoring, and within-pillar percentile ranking.
- **Aggregators** (two options): weighted arithmetic mean (fully compensatory: a surplus on one pillar offsets a deficit on another) and weighted geometric mean (partially compensatory, requires positive inputs; the positivity shift for z-scored columns is in Appendix B).

These are the standard alternatives from the composite-indicator literature (OECD/JRC, 2008). The model output for country i is its rank r_i(x).

### 4.3 Measures

**Rank shift and intervals.** For each prior we report the average absolute rank shift from Saisana et al. (2005): for every draw, we compute how many places each country moved from its published rank, then average across all countries and draws. Per country we also report the median simulated rank, interquartile range, and 90 percent rank interval (5th to 95th percentile). The width of the 90 percent interval is our per-country uncertainty measure. Global agreement with the published order is summarized by Spearman correlation (on scores) and Kendall correlation (on ranks).

**Sobol' indices.** A Sobol' sensitivity analysis attributes the total variance in a country's rank to each of the fourteen inputs and to combinations among them. For each input we report two indices:

- The **first-order index** S_j: the share of rank variance explained by input j on its own.
- The **total-effect index** S_Tj: the share of rank variance involving input j, either alone or through any interaction with other inputs.

Formal definitions and estimator details follow Sobol' (2001) and Saltelli et al. (2010) and appear in Appendix A. We use the Saltelli sampling design at base sample N = 1024, giving 16,384 model evaluations for our fourteen inputs, and estimate the indices with SALib (Herman and Usher, 2017). Because rank variance is highly unequal across countries (Denmark's rank barely moves; a mid-table country's rank moves

a lot), a simple average of per-country Sobol' indices would drown the informative countries in noise. We aggregate the country-level indices with weights proportional to each country's rank variance, so the aggregate answers the operative question: which inputs drive the rank movement that actually occurs?

**Realized importance.** Following Paruolo et al. (2013), we measure a pillar's realized influence on the composite by the **correlation ratio** $\eta^2_j$: the share of variance in the composite score that can be explained by knowing only pillar j's score. Intuitively, if $\eta^2_j$ is close to 1, pillar j is redundant with the rest of the index (knowing it tells you almost everything); if $\eta^2_j$ is closer to 0, pillar j carries information the other pillars do not. We estimate $\eta^2_j$ by binning pillar j's scores into 10 quantile bins and computing the share of composite variance explained by between-bin means (details in Appendix A). We then compare each pillar's realized share $\hat{\eta}^2_j / \Sigma \hat{\eta}^2_l$ against its assigned nominal share 1/12.

---

## 5. Results

All computation uses a single fixed seed; the full run completes in minutes on a laptop, and software versions are pinned in the repository.

### 5.1 Replication is exact

As stated in Section 3: maximum score deviation $5 \times 10^{-11}$, all 167 ranks reproduced. The pillar-to-index step of the Legatum index contains no undocumented processing.

### 5.2 The ranking is globally stable under every prior

Table 1 reports the weight-uncertainty results. The near-equal prior barely moves ranks ($\bar{R}_S = 0.97$; median $\tau = 0.987$). The grid prior shifts ranks 2.93 places on average. Even the uniform prior, which permits any pillar to dominate or vanish, yields $\bar{R}_S = 5.73$ with a median Spearman correlation of 0.987 against the published order and a 5th percentile of 0.965. Across 30,000 draws, no reweighting produces a wholesale disagreement with the published ranking; the ordering as a whole is a stable object. Figure 1 shows how $\bar{R}_S$ scales with the agnosticism of the prior.

**Table 1.** Weight-uncertainty analysis, 10,000 draws per prior. $\bar{R}_S$ is the average absolute rank shift; correlations are against the baseline on a 200-draw systematic subsample. Lower $\bar{R}_S$ and higher correlations indicate a more stable ranking.

| Weighting prior | $\bar{R}_S$ | Median $\rho$ | 5th pct. $\rho$ | Median $\tau$ | 5th pct. $\tau$ |
|---|---|---|---|---|---|
| Near-equal prior | 0.97 | 0.9995 | 0.9988 | 0.987 | 0.977 |
| Grid prior {0.5, 1, 1.5, 2} | 2.93 | 0.9965 | 0.9898 | 0.957 | 0.925 |
| Uniform prior (agnostic) | 5.73 | 0.9872 | 0.9654 | 0.912 | 0.855 |

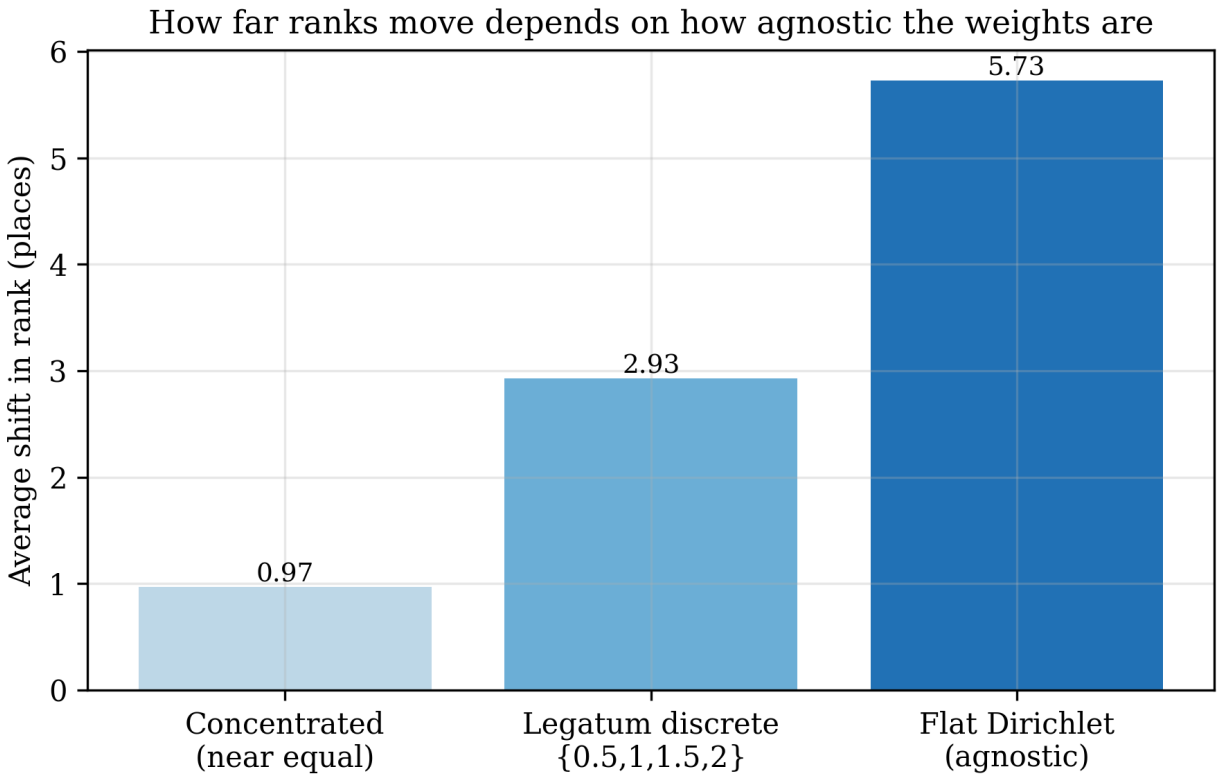


Figure 1: Figure 1

*Figure 1. Average absolute rank shift under the three priors.*

### 5.3 Individual mid-table positions are not

Figure 2 plots all 167 per-country 90 percent rank intervals under the uniform prior; the median width is 22 places. The extremes are pinned: Denmark (rank 1) occupies [1, 3] and South Sudan (rank 167) occupies [164, 167]. A country that leads or trails on most pillars does so under nearly any weighting. The middle of the table is different. Turkmenistan (published 107) spans [82, 137], Belarus (78) spans [53, 105], and China (54) spans [44, 93]. Full extremes appear in Appendix Table 5.

Figure 3 plots interval width against published rank, and Table 2 gives the tier profile: median widths of 8, 17, 37, 29.5, and 20 places across the five tiers, peaking for ranks 68–100. The mechanism is composition, not noise. Mid-table countries have heterogeneous pillar profiles, strong on some dimensions and weak on others, so their pairwise order is decided by which pillars the weights favor. Countries near the extremes are uniformly strong or weak, and no reweighting reorders them far. A reader who treats a published rank of 80 as meaningfully different from 95 is asserting a precision the index does not possess.

*Figure 2. Per-country rank uncertainty under the uniform prior, 10,000 draws. Band: 90 percent rank interval. Dots: median simulated rank. Ticks: published rank.*

*Figure 3. Interval width against published rank: uncertainty concentrates in the middle of the table.*

**Table 2.** Width of the 90 percent rank interval by tier of the published ranking, uniform prior.

| Published rank tier | Mean width | Median width | Max width |
|---|---|---|---|
| 1–33 | 9.9 | 8.0 | 28 |
| 34–67 | 20.0 | 17.0 | 49 |
| 68–100 | 35.6 | 37.0 | 52 |
| 101–134 | 30.4 | 29.5 | 55 |
| 135–167 | 20.2 | 20.0 | 45 |

### 5.4 Two weights dominate rank variance, and functional form is not a technicality

Figure 4 shows the variance-weighted Sobol' indices for all fourteen inputs (full table in Appendix Table 3). The weight on Personal Freedom carries S_T = 0.229 (S = 0.195) and Natural Environment S_T = 0.149 (S = 0.111); together they exceed the next five inputs combined. The aggregation selector, at S_T = 0.086, out-ranks nine of the twelve individual weights. The normalization selector reaches S_T = 0.068 against the smallest weight (S_T = 0.054, Investment Environment).

First-order indices sum to 0.889, so the rank map is close to additive in its inputs. The structure of the remaining interactions is informative: the weight inputs are dominated by main effects, whereas the normalization selector draws 65 percent of its

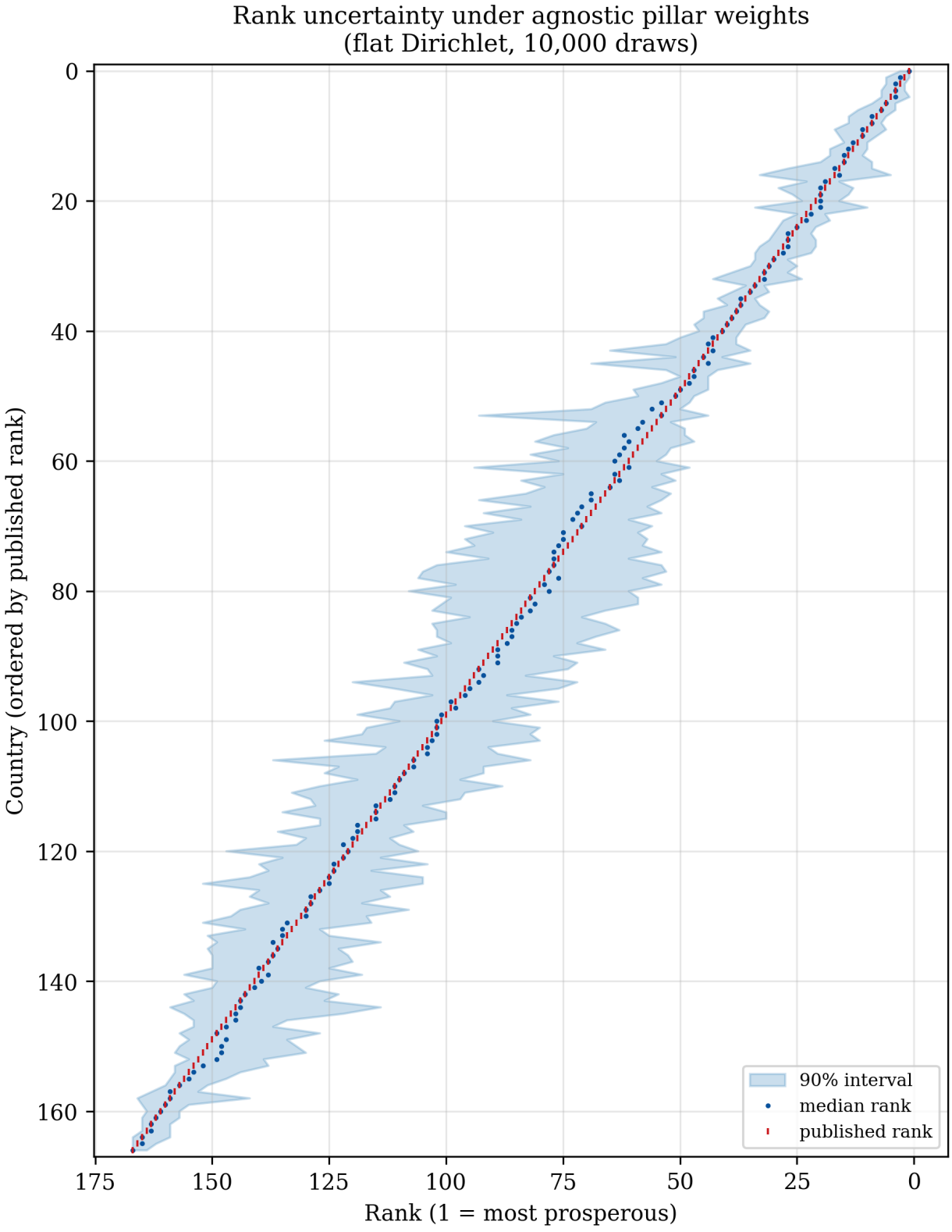


Figure 2: Figure 2

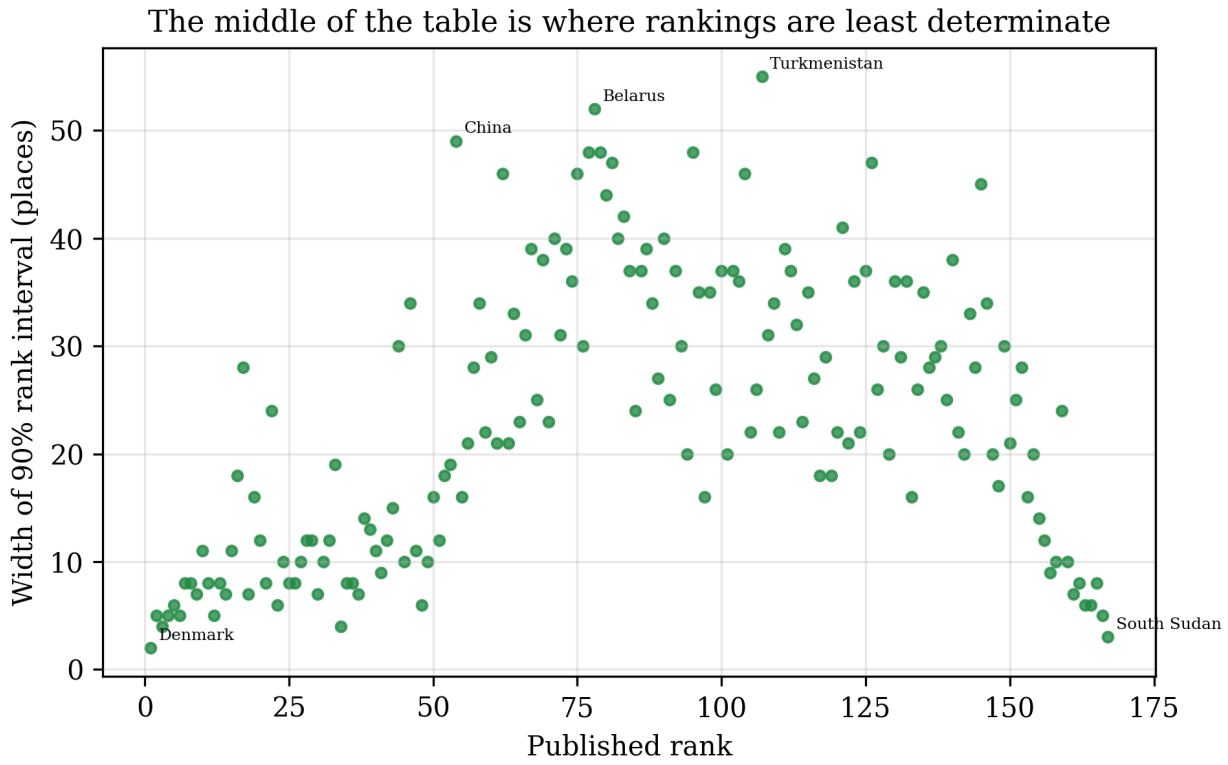


Figure 3: Figure 3

total effect from interactions (S = 0.024 against S_T = 0.068) and the aggregation selector 45 percent. This is the expected signature of functional-form choices, whose consequences depend on which weights are active. A sensitivity audit varying normalization alone, with weights fixed, would understate normalization's influence by roughly a factor of three.

*Figure 4. First-order (S_i) and total-effect (S_Ti) Sobol' indices for the fourteen inputs, aggregated across countries. Saltelli design, N = 1024, 16,384 evaluations.*

### 5.5 Leverage concentrates where correlation with the composite is weakest

Figure 5 compares nominal and realized importance (full table in Appendix Table 4). Realized shares run from 0.095 (Infrastructure and Market Access, $\hat{\eta}^2 = 0.90$) to 0.060 (Natural Environment, $\hat{\eta}^2 = 0.56$) against the uniform nominal 0.083. Personal Freedom ($\hat{\eta}^2 = 0.68$) and Social Capital ($\hat{\eta}^2 = 0.61$) join Natural Environment at the bottom.

Joining this to the Sobol' results gives the paper's central quantitative finding: across the twelve pillars, the total effect of a pillar's weight is strongly anticorrelated with the pillar's $\hat{\eta}^2$ (Spearman $-0.76$, $p = 0.004$; $-0.81$ for first-order indices). The mechanism is direct. A pillar with $\hat{\eta}^2$ near 0.9 is largely spanned by the consensus of the other eleven; perturbing its weight redistributes emphasis among near-duplicates, and ranks barely respond. A pillar with $\hat{\eta}^2$ near 0.56 carries information the rest of

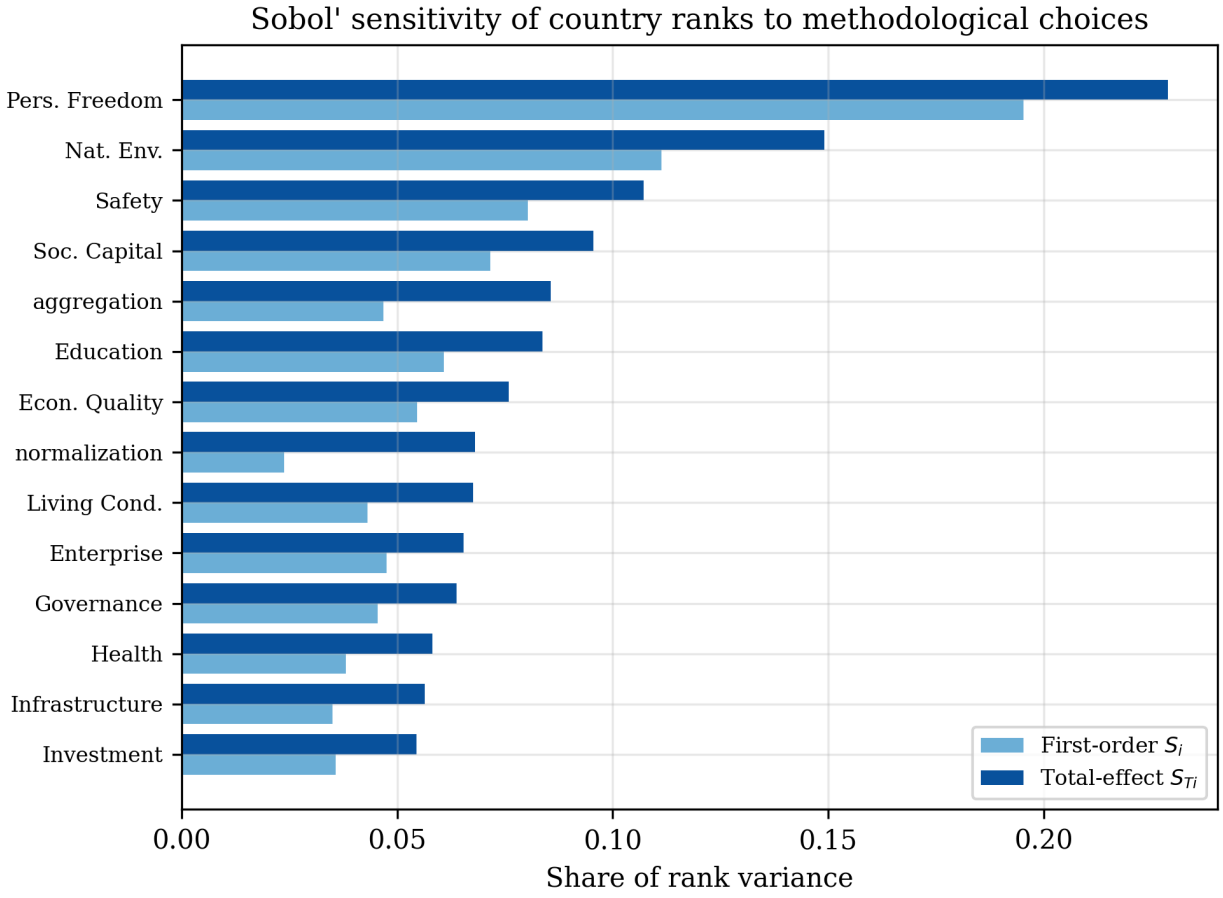


Figure 4: Figure 4

the index does not, and its weight decides how much that independent information counts. Equal weighting therefore operates as a substantive commitment: it grants environmental performance and civil liberty the same voice as the tightly clustered economic-institutional pillars, and it is exactly this commitment, not the many near-redundant weights, that the mid-table rankings are sensitive to.

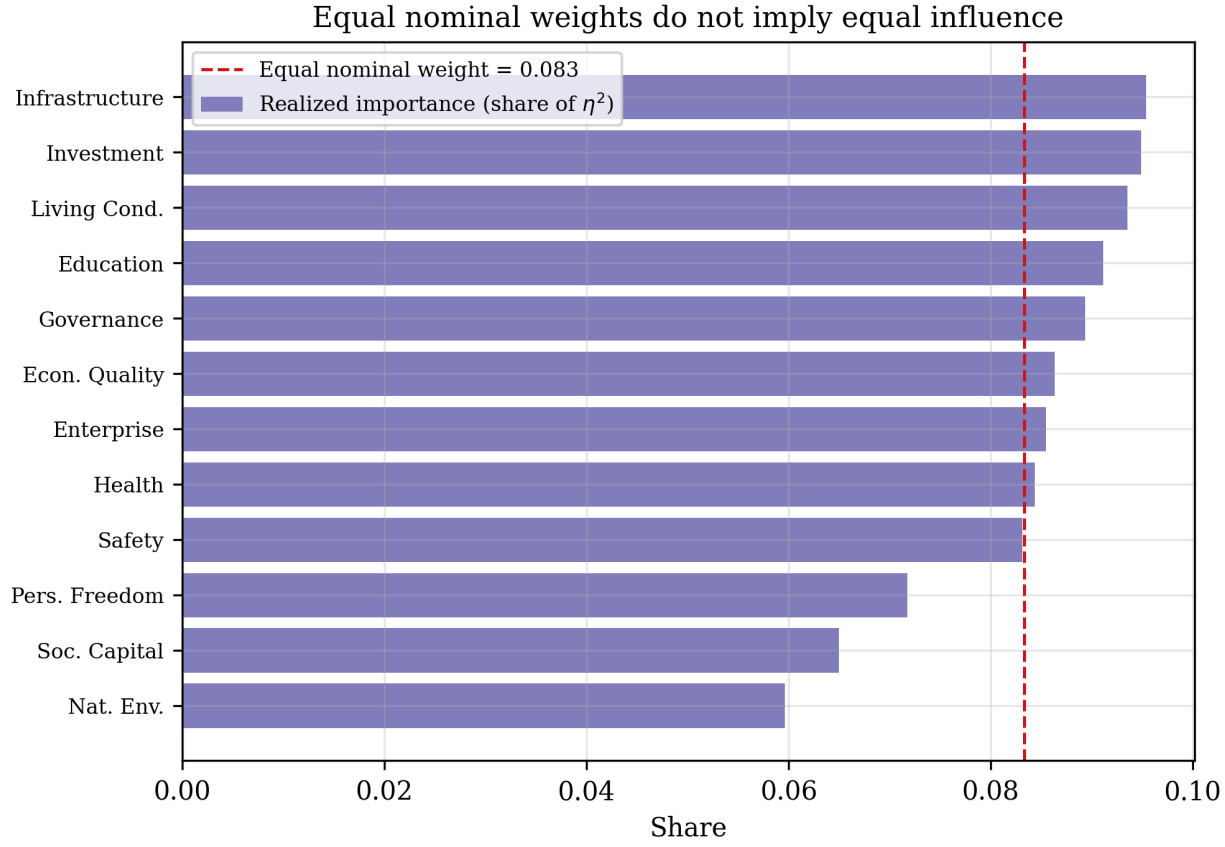


Figure 5: Figure 5

*Figure 5. Realized importance (share of main-effect η²) per pillar against the uniform nominal weight 1/12.*

---

## 6. Discussion

Three findings carry the practical weight. First, the index replicates exactly from public data, which is more than can be said of many published composites; independent auditing of this index is cheap, and the Prosperity Institute deserves credit for that. Second, mid-table point ranks overstate the precision of the measurement: a published rank near the middle is better read as a band of roughly ±15–20 places. Users should treat few-place differences between mid-table countries as noise, and any claim built on such differences should be checked against the grid-prior perturbation before being repeated. Third, the sensitivity structure localizes exactly where

the index is contestable. The near-redundant economic-institutional pillars could be reweighted substantially with little consequence. What the ranking actually depends on is how much voice the weakly correlated pillars, Personal Freedom, Natural Environment, and Social Capital, are given. Equal weighting resolves that question silently, in favor of full voice, and neither producers nor users typically recognize that a decision has been made.

The recommendation to producers follows from the cost asymmetry. The entire audit runs in minutes, so there is no computational excuse for publishing point ranks alone. Rank-uncertainty intervals generated as in Section 4.1, plus a one-page table of headline-rank sensitivity to the normalization and aggregation alternatives, would let readers see at a glance which comparisons the index can support. The audit transfers with trivial changes to any composite whose top-level aggregation is published, and the discipline it enforces, exact replication before perturbation, is worth making standard.

---

## 7. Limitations

We perturb only the pillar-to-index step. Indicator selection, imputation, the distance-to-frontier normalization of raw indicators, and all aggregation below the pillar level are held at published values, so our intervals are lower bounds on total structural uncertainty. The uniform prior is deliberately diffuse, and readers who find it too permissive should anchor on the grid-prior numbers ($\bar{R}\_S = 2.93$), which mirror how designers actually adjust weights. The geometric-aggregation branch requires positive inputs, and the positivity shift applied to z-scored columns (Appendix B) is one ad hoc repair among several. Finally, this is a single cross-section; repeating the audit across editions would separate durable structure from year-specific accident, and we leave that to future work.

---

## 8. Conclusion

We audited the 2023 Legatum Prosperity Index end to end: exact replication, Monte Carlo propagation of weight uncertainty under three priors, a fourteen-input global sensitivity analysis, and a quantified comparison of nominal weights with realized importance. The ranking is globally stable and locally soft, with a median 90 percent rank interval of 22 places that peaks at 55 in the middle of the table. Rank variance concentrates in the weights on Personal Freedom and Natural Environment together with the aggregation rule, and a weight's leverage is governed by its pillar's independence from the consensus (Spearman $-0.76$ between S_T and $\hat{\eta}^2$). Equal weighting is a substantive decision about how much that independent information counts. Composite rankings should ship with the uncertainty statements that make such decisions visible, and this paper releases the pipeline that produces them.

---

## Data and Code Availability

The Legatum dataset is distributed by the Prosperity Institute at https://index.prosperity.com and is not redistributed with our code, per its terms of use. All analysis code, precomputed outputs, figures, and reproducibility tests (fixed seed; pytest suite verifying exact replication and the stability of the headline Monte Carlo statistic) are available at https://github.com/RE17-IWD/legatum-prosperity-robustness.

---

## References


[1] W. Becker, M. Saisana, P. Paruolo, and I. Vandecasteele. Weights and importance in composite indicators: Closing the gap. *Ecological Indicators*, 80:12–22, 2017.

[2] J. Herman and W. Usher. SALib: An open-source Python library for sensitivity analysis. *Journal of Open Source Software*, 2(9):97, 2017.

[3] Legatum Institute. *The Legatum Prosperity Index 2023: Methodology Report*. Legatum Institute Foundation, London, 2023.

[4] OECD and JRC European Commission. *Handbook on Constructing Composite Indicators: Methodology and User Guide*. OECD Publishing, Paris, 2008.

[5] P. Paruolo, M. Saisana, and A. Saltelli. Ratings and rankings: voodoo or science? *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(3):609–634, 2013.

[6] M. Saisana, A. Saltelli, and S. Tarantola. Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168(2):307–323, 2005.

[7] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, 181(2):259–270, 2010.

[8] I. M. Sobol'. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55(1–3):271–280, 2001.


---

## Appendix A. Formal Definitions and Distributional Details

**Average absolute rank shift (Section 4.3).** For M = 10,000 draws and n = 167 countries,

$$\bar{R}_S = \frac{1}{nM} \sum_{i=1}^{n} \sum_{m=1}^{M} \left| r_i(w^{(m)}) - r_i \right| .$$

**Sobol' indices (Section 4.3).** For rank output Y = r_i(X) with independent inputs X = (X_1, ..., X_{k'}), the variance decomposes as Var(Y) = Σ_j V_j + Σ_{j<l} V_{jl} + ⋯ with V_j = Var[E(Y | X_j)] (Sobol', 2001). The first-order and total-effect indices are

$$S_j = \frac{V_j}{\mathrm{Var}(Y)}, \qquad S_{Tj} = 1 - \frac{\mathrm{Var}[E(Y|X_{-j})]}{\mathrm{Var}(Y)},$$

where X_{−j} denotes all inputs other than X_j. The variance-weighted aggregation across countries is

$$\mathcal{S}_j = \sum_i \pi_i S_j^{(i)}, \qquad \pi_i = \frac{\mathrm{Var}(r_i(X))}{\sum_l \mathrm{Var}(r_l(X))}.$$

**Correlation ratio (Section 4.3).** For pillar j with score column s_j, cut into B = 10 quantile bins G_1, ..., G_B, the estimator is

$$\hat{\eta}_j^2 = \frac{\sum_{b=1}^{B} |G_b|(\bar{y}_{G_b} - \bar{y})^2}{\sum_i (y_i - \bar{y})^2}.$$

The estimator makes no functional-form assumption, and results are insensitive to reasonable B.

**Uniform prior (Section 4.1).** The flat Dirichlet Dir(1_k) has density Γ(k), constant on the simplex Δ^(k−1), so it is the uniform distribution on the space of valid weight vectors.

**Hypercube-to-simplex map (Section 4.2).** For Saltelli sampling we generate weights on the simplex from independent uniform inputs by the inverse-exponential map

$$w_j(u) = \frac{-\ln u_j}{\sum_{l=1}^{k} (-\ln u_l)}.$$

If U_j i.i.d. ~ Uniform(0, 1) then −ln U_j i.i.d. ~ Exponential(1), and the normalized vector of k independent Exponential(1) draws is exactly Dir(1_k). The map therefore carries the Saltelli hypercube onto the uniform simplex without breaking the independent-uniform input geometry the estimators assume.

**Near-equal prior moments (Section 4.1).** For w ~ Dir(c · 1_k), E[w_j] = 1/k and Var(w_j) = (k − 1) / (k² (ck + 1)). With c = 50 and k = 12, sd(w_j) ≈ 0.011 around 1/12 ≈ 0.083.

---

## Appendix B. Normalization and Aggregation Alternatives

Normalizations apply column-wise across the n countries. Let $\bar{s}_j$, $\sigma_j$, $s^{\min}_j$, $s^{\max}_j$ denote the column mean, standard deviation, minimum, and maximum.

- **Identity.** $N_1(s_{ij}) = s_{ij}$, the native distance-to-frontier scale.
- **Min–max.** $N_2(s_{ij}) = 100\,(s_{ij} - s^{\min}_j) / (s^{\max}_j - s^{\min}_j)$, forcing each pillar onto the full [0, 100] range regardless of observed spread.
- **z-score.** $N_3(s_{ij}) = (s_{ij} - \bar{s}_j) / \sigma_j$, equalizing pillar variances and admitting negative values.
- **Percentile rank.** $N_4(s_{ij}) = (100 / n) \cdot \#\{l : s_{lj} \le s_{ij}\}$, retaining only within-pillar order.
- **Weighted arithmetic mean.** $A_1(m_i; w) = \Sigma_j w_j m_{ij}$: fully compensatory, a deficit on one pillar offsets one-for-one against a surplus on another.
- **Weighted geometric mean.** $A_2(m_i; w) = \exp(\Sigma_j w_j \ln(m_{ij} + \delta_j))$: partially compensatory, defined only for positive arguments. The shift $\delta_j = \max(0, -\min_i m_{ij}) + 10^{-6}$ is applied to any column with non-positive entries, which arise under $N_3$.

---

## Appendix C. Additional Results

**Table 3.** Variance-weighted Sobol' indices, all fourteen inputs, sorted by total effect. First-order indices sum to 0.889.

| Input | $S_i$ | $S_{Ti}$ |
|---|---|---|
| Weight: Personal Freedom | 0.195 | 0.229 |
| Weight: Natural Environment | 0.111 | 0.149 |
| Weight: Safety and Security | 0.080 | 0.107 |
| Weight: Social Capital | 0.072 | 0.096 |
| Aggregation rule | 0.047 | 0.086 |
| Weight: Education | 0.061 | 0.084 |
| Weight: Economic Quality | 0.055 | 0.076 |
| Normalization rule | 0.024 | 0.068 |
| Weight: Living Conditions | 0.043 | 0.068 |
| Weight: Enterprise Conditions | 0.048 | 0.065 |
| Weight: Governance | 0.045 | 0.064 |
| Weight: Health | 0.038 | 0.058 |
| Weight: Infrastructure and Market Access | 0.035 | 0.056 |
| Weight: Investment Environment | 0.036 | 0.054 |

**Table 4.** Nominal weight versus realized importance for all twelve pillars, sorted by importance share. The nominal share is 1/12 ≈ 0.083 throughout.

| Pillar | Nominal share | $\hat{\eta}^2$ | Importance share |
|---|---|---|---|
| Infrastructure and Market Access | 0.083 | 0.899 | 0.095 |

| Pillar | Nominal share | $\hat{\eta}^2$ | Importance share |
|---|---|---|---|
| Investment Environment | 0.083 | 0.894 | 0.095 |
| Living Conditions | 0.083 | 0.881 | 0.094 |
| Education | 0.083 | 0.859 | 0.091 |
| Governance | 0.083 | 0.842 | 0.089 |
| Economic Quality | 0.083 | 0.814 | 0.086 |
| Enterprise Conditions | 0.083 | 0.806 | 0.086 |
| Health | 0.083 | 0.795 | 0.084 |
| Safety and Security | 0.083 | 0.783 | 0.083 |
| Personal Freedom | 0.083 | 0.676 | 0.072 |
| Social Capital | 0.083 | 0.612 | 0.065 |
| Natural Environment | 0.083 | 0.562 | 0.060 |

**Table 5.** The six widest 90 percent rank intervals under the uniform prior, with the pinned extremes for contrast.

| Country | Published rank | 5th pct. | 95th pct. | Width |
|---|---|---|---|---|
| Turkmenistan | 107 | 82 | 137 | 55 |
| Belarus | 78 | 53 | 105 | 52 |
| China | 54 | 44 | 93 | 49 |
| Russia | 77 | 54 | 102 | 48 |
| Saudi Arabia | 79 | 58 | 106 | 48 |
| Turkey | 95 | 72 | 120 | 48 |
| Denmark | 1 | 1 | 3 | 2 |
| South Sudan | 167 | 164 | 167 | 3 |