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, USA

Abstract

The Legatum Prosperity Index publishes a ranking of 167 countries according to an equal-weighted mean of twelve scores that show each country's performance across economic and social "pillars." Here we audit the degree to which that ranking survives if the equal weights of $1/12$ are allowed to vary, together with the choices of normalization and aggregation the construction of the index requires. The equal-weighted mean of the published pillar scores is found to reproduce each of the official scores to within 5×10^{-11} and each of the official country rankings exactly, so that any rank shift in the results arises from our perturbations and not from reconstruction error. The resulting ranking remains globally stable when the twelve weights are allowed to vary according to three statistical priors of increasing agnosticism (from a distribution tightly concentrated around $1/12$ to a fully uniform draw over the simplex), and also when the twelve weights vary together with two selection variables governing normalization and aggregation according to a variance-based sensitivity analysis of 16,384 model evaluations. The ranking is globally stable in that the median Spearman rank correlation between the original and the perturbed rankings stays at or above 0.987 under each of the three priors; however, the ranking is locally soft, as the median 90-percentile rank interval reaches 22 places under the fully agnostic prior, increasing to a median of 37 places for the ranks of 68 to 100 and to a maximum of 55 places for Turkmenistan. The Sobol' sensitivity indices indicate that the weights exhibiting the strongest total effect on the ranking are those for the pillars of Personal Freedom ($S_T = 0.23$) and Natural Environment ($S_T = 0.15$); across the twelve pillars, the total effect of each pillar's weight is found to be strongly anticorrelated with that pillar's main-effect η^2 against the composite Legatum Prosperity Index score (Spearman rank correlation -0.76, $p = 0.004$), so that the weights whose pillars most correlate with the composite score have the weakest total effects upon it. Finally, the aggregation selection variable is found to have a more powerful total effect than nine of the twelve individual pillar weights, and the normalization selection variable draws 65 percent of its total effect from interactions with the pillar weights. Thus, within the framework of this model, the equal weighting of the twelve pillars is a substantive choice regarding the relative information content of the environmental and civil-liberty pillars, rather than a neutral default value.

Keywords: composite indicators, Legatum Prosperity Index, Monte Carlo simulation, Sobol' indices, sensitivity analysis, rank robustness, country rankings, uncertainty quantification, variance-based sensitivity analysis, correlation ratio, weighting schemes, reproducibility, Saltelli sampling

1. Introduction

Composite indices reduce hundreds of indicators for each country to a single number. That number enters policy documents, the lexicon of journalism, and economic regressions. Every such index embeds three construction choices that the readers of its reports 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 the ranks of the countries represented in the index enough to reverse the conclusions drawn from that index. The OECD–JRC handbook on the construction of composite indices later recommended that a sensitivity analysis of these three choices be included in the construction of any composite index (OECD/JRC, 2008). The producers of the indices rarely include such an analysis in their reports.

We provide such an analysis for the 2023 edition of the Legatum Prosperity Index. There are three reasons why the Legatum index is a good candidate for audit. First, it is cited widely in the media and in academic work. Second, it has reported on 167 countries every year since 2007. Third, the construction of the index is exactly documented by its authors: the equal-weighted arithmetic mean of the scores of twelve "pillars" of the index (Legatum Institute, 2023). This last property is what makes an audit of its construction possible at all, because it permits exact numerical replication of the published rankings before any perturbation is introduced. Finally, the equal weighting of the pillars raises a question as to whether the nominal equal weight of $1/12$ for each pillar reflects the effective importance of that pillar in determining the rank of each country. It is known from studies of the construction of composite indices that the nominal weight of a component is not necessarily the same as its effective weight, or importance, in the composite (Paruolo et al., 2013; Becker et al., 2017). Whether the difference between the nominal and effective weights is material for the Legatum index is an empirical question.

Four findings emerge from the audit. First, reproducing the calculation of the Legatum index as published by its authors reproduces the rankings of all 167 countries with a maximum deviation in the scores of 5×10^{-11}; this very small deviation indicates that there is no undocumented processing of the pillar scores before the formation of the index score. Second, a Monte Carlo simulation of the twelve weights of the pillars of the index under three different priors of increasing agnosticism yields, for each country, 90-percent rank intervals with a median width of 22 rank places under the fully agnostic prior; this width indicates the instability of the ranks of countries against perturbations in the weights of the pillars. Third, a Sobol' sensitivity analysis on the rank of countries as a function of the twelve weights of the pillars and the two selection variables governing normalization and aggregation reveals that the ranks of countries are most influenced by the weights of the pillars of Personal Freedom ($S_T = 0.23$) and Natural Environment ($S_T = 0.15$), and that the aggregation selection variable exhibits a stronger total effect than nine of the twelve individual pillar weights. Fourth, the Sobol' total effect of the weight of a pillar is strongly anticorrelated with the main-effect η^2 of that pillar's score against the composite Legatum Prosperity Index score (Spearman rank correlation -0.76, $p = 0.004$): the weights of the pillars whose scores most correlate with the composite score have the weakest total effects on that score, so that the weights that move the ranking most protect the information content of the pillars least correlated with the rest of the index.

---

2. Related Work

Uncertainty and sensitivity analysis of composites. This methodological template was developed by Saisana et al. (2005). The uncertainty in the normalization, weighting, and aggregation steps of the composite construction process is propagated through Monte Carlo sampling, and the resulting distribution of composite ranks is decomposed into components using global sensitivity indices. This approach is described in the OECD–JRC handbook on the construction of composite indicators (OECD/JRC, 2008), which also describes the standard alternatives at each of the three steps. Despite the availability of this tool, sensitivity analyses have not yet been performed on many of the widely cited country rankings currently in use, and the producers of these rankings do not make available any uncertainty statements with their results.

Variance-based sensitivity analysis. The functional-ANOVA approach to sensitivity analysis originates in the work of Sobol' (2001). The sampling designs and estimators that allow for calculation of the first-order and total-effect sensitivity indices require only $N(k+2)$ model evaluations (Saltelli et al., 2010); these methods have been implemented in the Python library SALib by Herman and Usher (2017), which we use in this paper. Although the model output is not a continuous quantity but the discrete rank of a country in the ranking, the functional-ANOVA decomposition applies without modification because it requires only that the model output be square-integrable.

Weights versus importance. Paruolo et al. (2013) show that the nominal weight assigned to each component in the construction of a composite is a poor measure of the effective importance of that component. They propose the main-effect correlation ratio η^2 of a component's score against the composite as the appropriate importance measure. Alternative methods for closing the gap between assigned weights and realized importance have been proposed by Becker et al. (2017). We use the correlation-based importance measure, but we add an observation that is quantified here for the Legatum index: the leverage of a weight over the ranking is itself governed by the η^2 of the corresponding pillar, so that leverage concentrates where correlation with the composite is weakest.

---

3. Data and Baseline Replication

The Legatum Prosperity Index calculates a score for 167 countries via a three-level architecture. Three hundred indicators are normalized through a distance-to-frontier transformation and combined into 67 elements, which are then combined into twelve pillars of the index, each on a 0–100 scale. The twelve pillars cover the following categories: 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. The pillar scores are then combined by an equal-weighted arithmetic mean to produce the overall index score for each country (Legatum Institute, 2023). This published table of pillar scores is used as the data for our analysis, and only the step from pillar scores to the overall index score is perturbed; Section 7 discusses this scoping choice.

Under a weight vector $w = (w_1, \dots, w_{12})$ whose entries are non-negative and sum to one, the composite score for country i is the sum, across pillars, of that country's pillar scores multiplied by the corresponding weights, and the rank of country i is one plus the number of countries with a higher composite score. The published index corresponds to the case in which every weight equals $1/12$.

Computed from the published pillar-score data file, the equal-weighted arithmetic mean reproduces the overall index score for each country with a maximum absolute deviation of 5×10^{-11}, and reproduces the rank of every country without exception (Spearman and Kendall correlations of exactly 1.0 over the 167 positions). Any rank interval reported below is therefore free of the reconstruction error that would otherwise confound the perturbation analysis.

---

4. Materials and Methods

4.1 Weight-uncertainty propagation

The first exercise holds the normalization and aggregation choices fixed and varies only the weights. Such an exercise answers the most commonly asked question about the equal-weighted index: do the weights matter? We bracket the space of plausible answers with three priors and draw $M = 10{,}000$ weight vectors from each.

The first of the three priors is the **uniform prior**. It selects weight vectors at random from the set of all twelve-dimensional weight vectors whose components are non-negative and sum to 1. Any one of the twelve pillars may therefore dominate the index or have no impact whatsoever on the calculated score for a given country. This uniform prior is equivalent to the flat Dirichlet distribution (Appendix A).

The second is the **near-equal prior**. It assigns each of the twelve weights to a value close to $1/12$, with a small standard deviation of approximately 0.011, so that no individual weight ever strays far from $1/12$. This prior reflects the position of an analyst who accepts equal weighting in general but does not believe in its exactness. It is realized with a concentrated Dirichlet distribution (Appendix A).

The third and final prior is the **grid prior**. It assigns each of the twelve weights independently to one of the values $\{0.5, 1, 1.5, 2\}$, then rescales the twelve components so that they sum to 1. This prior is selected because it mimics the way index designers typically adjust the weights away from equal, and is therefore the most policy-relevant of the three priors.

4.2 Joint perturbation design

The second exercise allows all three construction choices to be varied simultaneously. The Sobol' sampling design requires that the model inputs form a unit hypercube, so we work with fourteen coordinates in $[0, 1]$: twelve for the pillar weights and two additional coordinates that determine the normalization and aggregation methods. The twelve weight coordinates are transformed into a weight vector whose components sum to 1 through a standard map that reproduces the uniform prior of 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 used by Legatum
- Min–max normalization to the $[0, 100]$ range
- Z-scoring (standard scoring)
- Within-pillar percentile rank

Aggregators (two options):

- Weighted arithmetic mean (fully compensatory: a surplus in one pillar can be balanced out by a deficiency in another)
- Weighted geometric mean (partially compensatory: a surplus in one pillar is only partly balanced by a deficiency in another, and every pillar score must be positive; see Appendix B)

These are the standard alternatives from the composite-indicator literature (OECD/JRC, 2008). For each country in the sample, the model output is the rank $r_i(x)$ of that country under the perturbed choices.

4.3 Measures

Rank shift and intervals. For each of the three priors, we report the average absolute rank shift from Saisana et al. (2005): the average number of places a country moved from its published rank, taken across all countries and all draws. For each country we also report the median simulated rank, the interquartile range, and the 90-percent rank interval (the interval from the 5th to the 95th percentile of the simulated ranks). The width of the 90-percent rank interval is our measure of the uncertainty in the rank of each country. Finally, the Spearman and Kendall correlations between the published ranks and the perturbed ranks provide a measure of the agreement of the perturbed ranking with the published one.

Sobol' indices. For each of the fourteen inputs, the Sobol' indices measure the contribution of that input (and its interactions with other inputs) to the total variance in the rank of a country. For each input, two indices are provided:

- The **first-order index** S_j measures the fraction of the variance in the country's rank that is due to input j alone.
- The **total-effect index** S_{Tj} measures the fraction of the variance in the country's rank that is due to input j, either alone or in interaction with any of the other inputs.

The formal definitions and estimator details for these indices appear in Appendix A. The indices are estimated using SALib (Herman and Usher, 2017) at a Saltelli base sample size of $N = 1024$, corresponding to $N(k' + 2) = 16{,}384$ model evaluations for the $k' = 14$ inputs. Since the variance in the rank of each country is not the same (Denmark's rank barely moves, whereas the rank of a country in the middle of the distribution moves considerably), a simple average of the per-country Sobol' indices would mask the importance of some of the pillars for certain countries. We therefore aggregate the per-country indices with weights that are proportional to the rank variance of each country, so that the aggregate indices reflect which inputs actually drive the rank movement in the perturbations.

Realized importance. As in Paruolo et al. (2013), the realized importance of each pillar is measured by its correlation ratio η_j^2, which is the fraction of the variance in the composite score that is explained by that pillar's scores alone. A high value of η_j^2 indicates that pillar j is redundant with the rest of the index (its score is largely determined by the scores of the other eleven pillars), while a low value indicates that pillar j contributes information about the composite that is not carried by any of the other pillars. Further details on the estimation of η_j^2 appear in Appendix A. Finally, each pillar's realized importance share, $\hat{\eta}_j^2 / \sum_l \hat{\eta}_l^2$, is compared to the pillar's nominal share of $1/12$ in the equal-weighted index.

---

5. Results

All analyses used the same seed for random number generation, so that the results can be replicated on demand. The complete analysis takes only a few minutes to run on a laptop, and the code is made available in the repository accompanying this report.

5.1 Replication is exact

As stated in Section 3, the maximum absolute deviation between the composite scores computed from the published pillar-score data file and the officially published Legatum scores is 5×10^{-11}, and every one of the 167 official country ranks is reproduced without exception. The pillar-to-index step of the Legatum index therefore 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 τ = 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 stable. 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 ρ	5th pct. ρ	Median τ	5th pct. τ
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

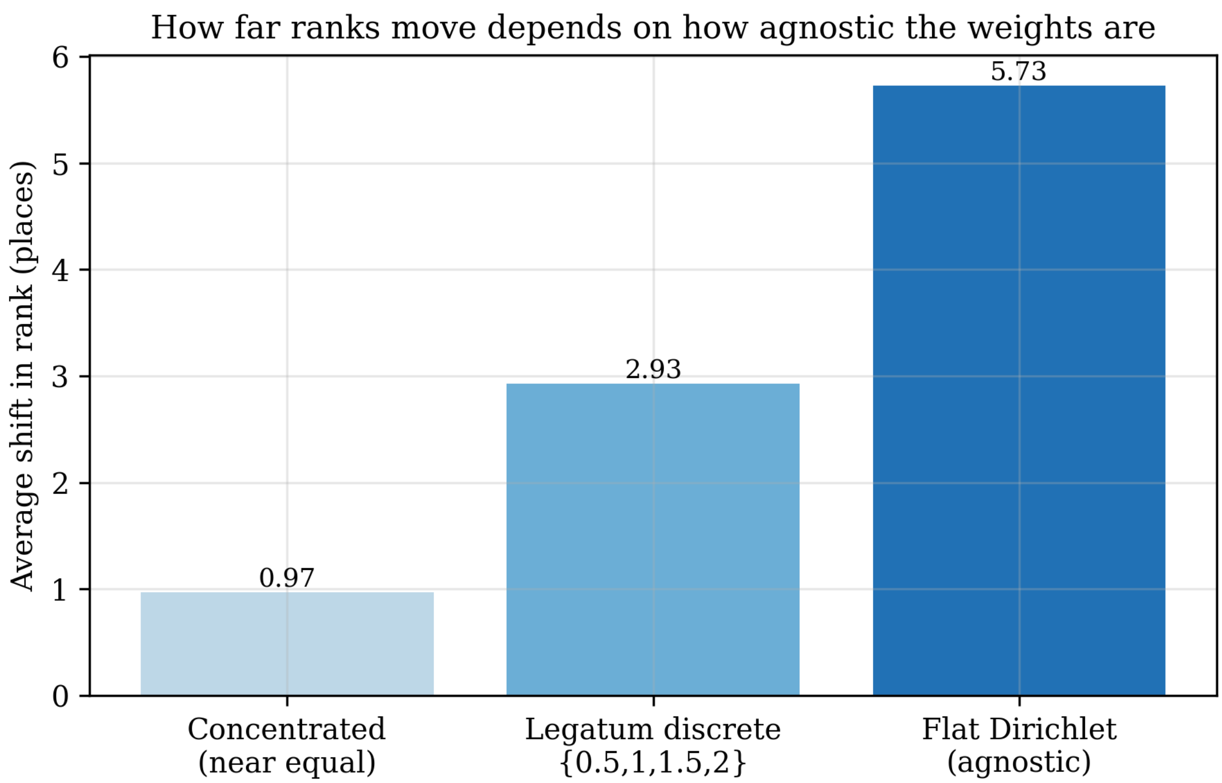

Figure 1: Figure 1

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 of these intervals is 22 places. The extremes are pinned: Denmark, with a published rank of 1, has a 90-percent interval of $[1, 3]$, while South Sudan, with a published rank of 167, has a 90-percent interval of $[164, 167]$. A country that leads or trails on most of the pillars of the index has that outcome under nearly any weighting of the pillars. The countries in the middle of the table have more heterogeneous intervals: Turkmenistan, whose published rank is 107, has a 90-percent interval of $[82, 137]$; Belarus, with a published rank of 78, has a 90-percent interval of $[53, 105]$; and China, at a published rank of 54, has a 90-percent interval of $[44, 93]$. The complete data are given in Appendix Table 5.

Figure 3 plots the width of these 90-percent intervals against each country's published rank. Table 2 gives the tier profile of the index: the median widths of the 90-percent intervals are 8, 17, 37, 29.5, and 20 places for the five rank tiers, respectively, and the widths peak for countries with published ranks between 68 and 100. For a country in the middle of the rank distribution, with relatively strong performances in some pillars and relatively weak performances in others, the rank of that country relative to other countries is determined by the relative weights of those pillars. For a country that is uniformly strong or uniformly weak across all pillars, however, there is no reweighting that reorders that country substantially relative to others. Thus, a reader who treats a country with a published rank of 80 as meaningfully differently ranked from a country with a published rank of 95 is asserting a precision of the rank that the index does not exhibit.

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 presents the variance-weighted Sobol' indices for all fourteen inputs to the model (full table in Appendix Table 3). The weight on the Personal Freedom pillar has the highest variance-weighted total-effect index of all individual inputs at $S_T = 0.229$ (first-order index $S = 0.195$), followed by the weight on the Natural Environment pillar at $S_T = 0.149$ ($S = 0.111$); together these two inputs account for more of the rank variance than the next five most influential inputs combined. Both the aggregation and normalization selectors out-rank the majority of individual pillar weights: the aggregation selector reaches $S_T = 0.086$, which exceeds nine of the twelve individual pillar weights, and the normalization selector reaches $S_T = 0.068$, compared to the smallest weight for any individual input ($S_T = 0.054$, for the Investment Environment pillar).

The first-order Sobol' indices for the fourteen inputs sum to 0.889, so the rank map is close to additive in its inputs. Furthermore, the pillar-weight inputs are dominated by their main effects, whereas the two selector inputs draw much of their total effect from interactions with other inputs: for the normalization selector, the first-order index is $S = 0.024$ against a total-effect index of $S_T = 0.068$, so that interactions

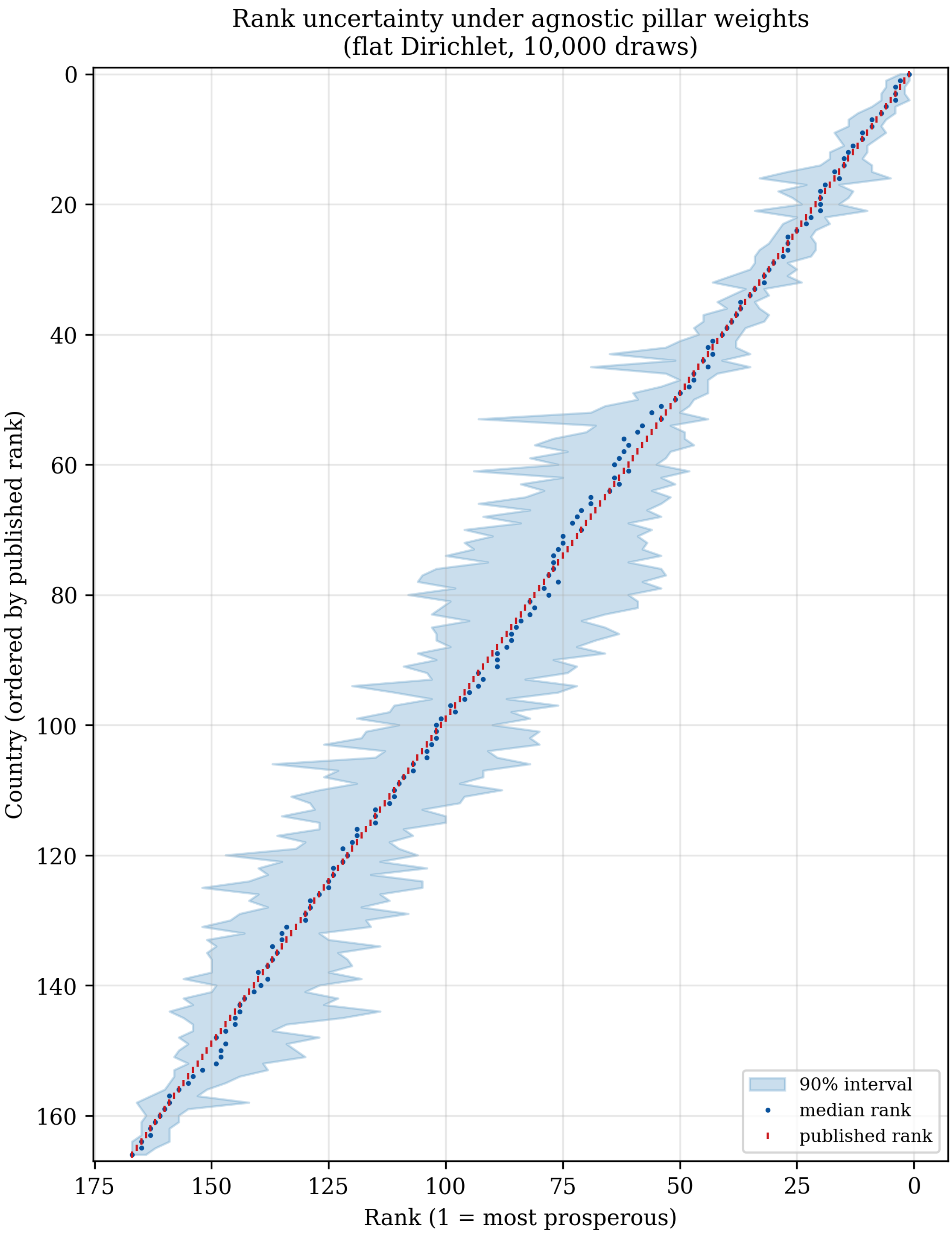

Figure 2: Figure 2

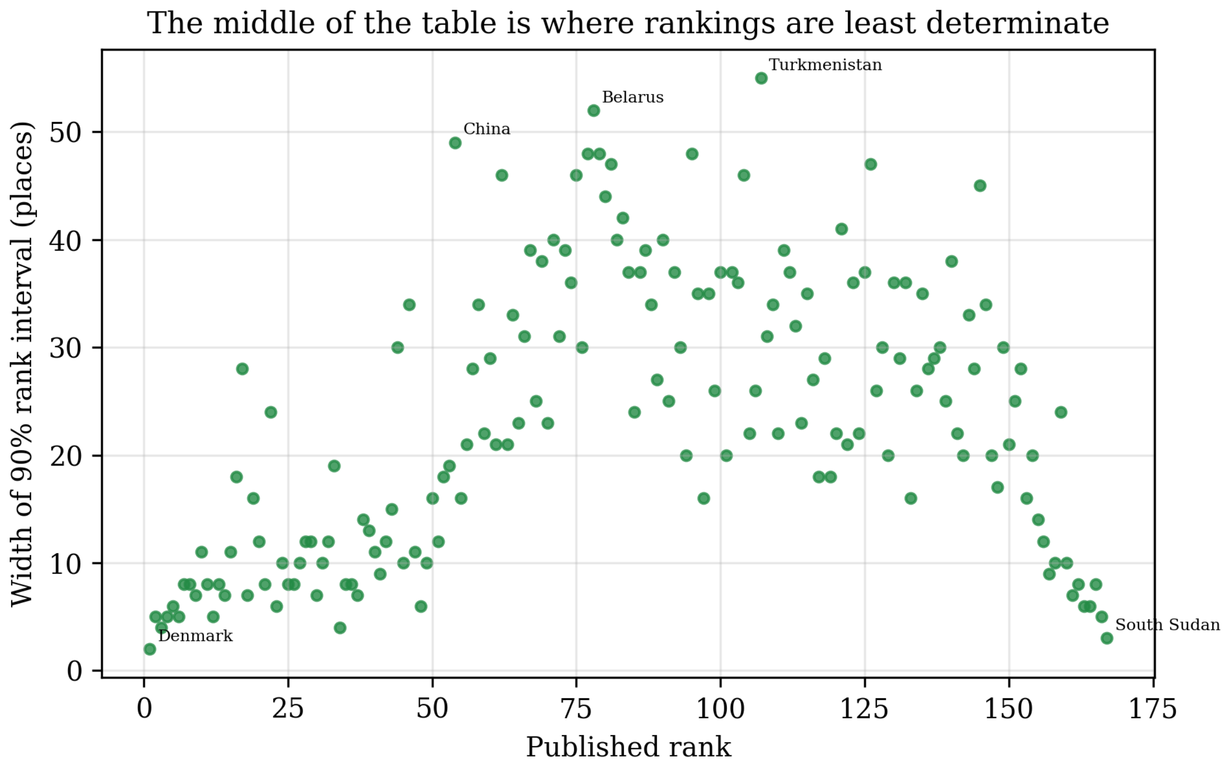

Figure 3: Figure 3

contribute 65 percent of the total effect; for the aggregation selector, interactions contribute 45 percent of the total effect. These interaction fractions are the expected signature of functional-form choices, since the consequences of a given normalization or aggregation rule depend on the specific weights against which it is being applied. Consequently, auditing the model's variance by varying only the normalization selector while holding the weights fixed would understate the influence of the normalization selector by a factor of approximately 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 plots the nominal and realized importance of each of the twelve pillars of the Legatum Prosperity Index (full table in Appendix Table 4). The realized importance shares range from a high of 0.095 for the Infrastructure and Market Access pillar ($\hat{\eta}^2 = 0.90$) to a low of 0.060 for the Natural Environment pillar ($\hat{\eta}^2 = 0.56$), against the uniform nominal share of 0.083. The shares for the remaining pillars fall between these two extremes. The Personal Freedom pillar ($\hat{\eta}^2 = 0.68$) and the Social Capital pillar ($\hat{\eta}^2 = 0.61$) join Natural Environment at the low end of the scale.

These realized-importance findings, taken together with the Sobol' results, yield the central quantitative finding of the paper: across the twelve pillars, the total effect of

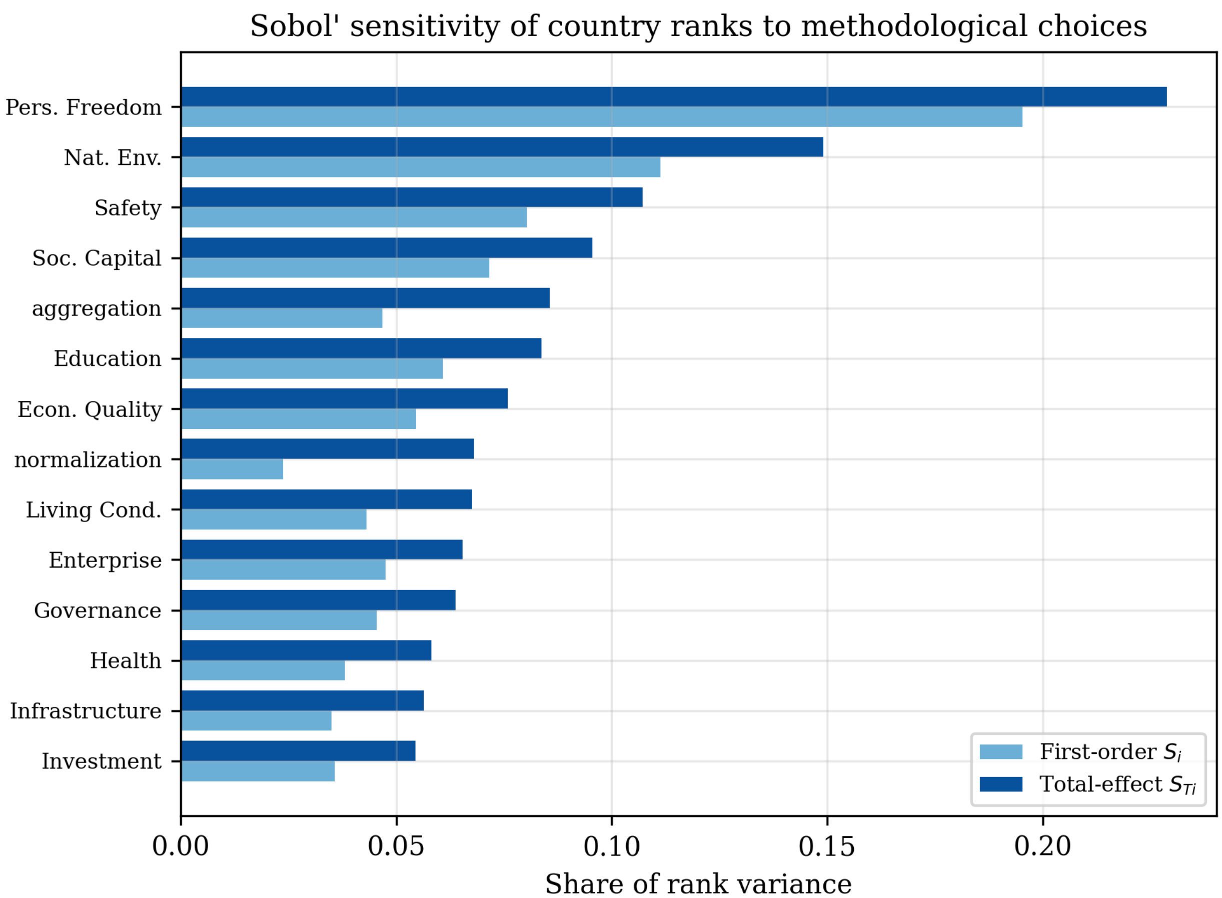

Figure 4: Figure 4

a pillar's weight is strongly anticorrelated with that pillar's $\widehat{\eta}^2$ against the composite Legatum Prosperity Index score (Spearman's rank correlation coefficient -0.76, $p = 0.004$; -0.81 for the first-order sensitivity indices). A pillar with a high value of $\widehat{\eta}^2$ is one whose score is largely explained by the consensus among the other eleven pillars; perturbing its weight redistributes emphasis among near-duplicates, and the composite score changes very little. A pillar with a low $\widehat{\eta}^2$, such as Natural Environment, carries information about the composite that is not contained in the other pillars; its weight therefore has a large effect on the composite score, and on the ranks that composite score produces. The equal weighting of the twelve pillars is therefore a substantive commitment: it grants environmental performance and civil liberty the same voice in the composite as the tightly clustered economic and institutional pillars, and it is exactly this substantive commitment, rather than the many near-redundant pillar weights, that the mid-table rankings are sensitive to.

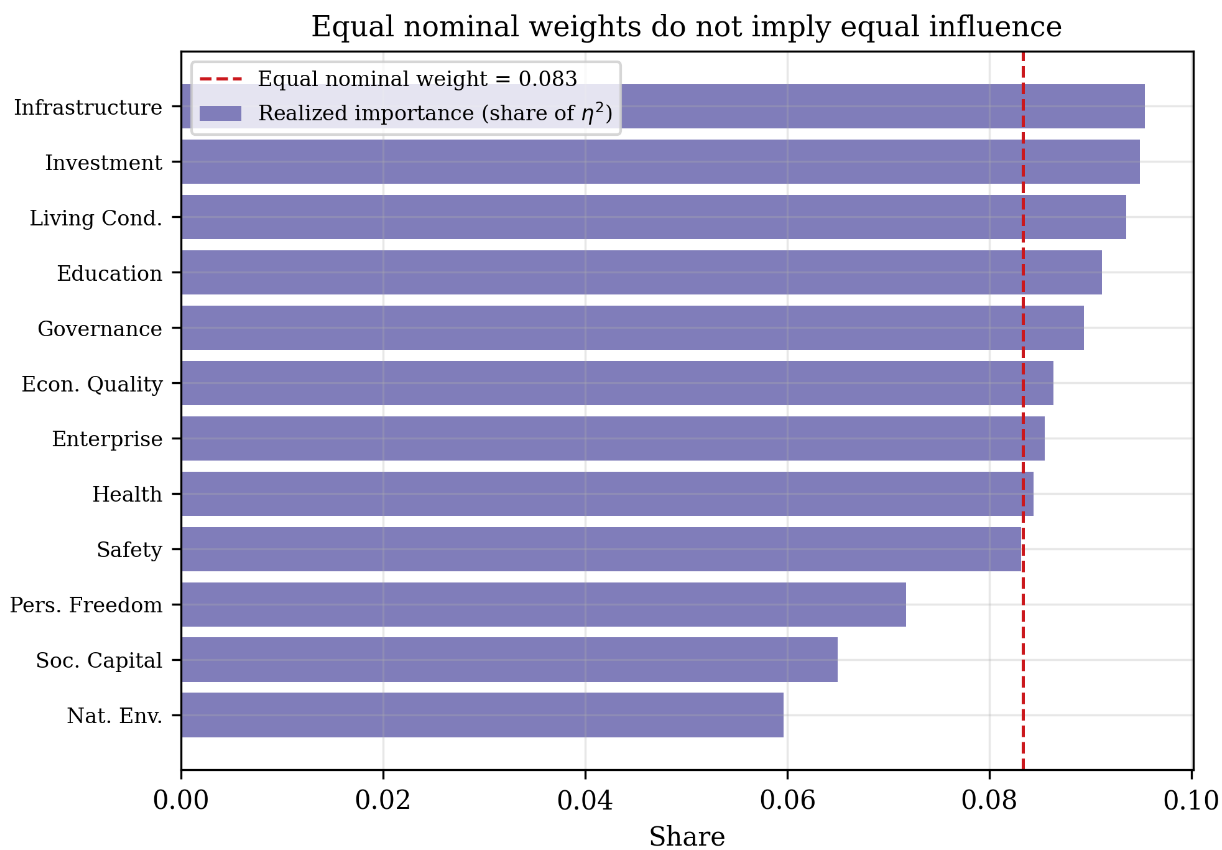

Figure 5: Figure 5

Figure 5. Realized importance (share of main-effect η^2) per pillar against the uniform nominal weight 1/12.

---

6. Discussion

The index replicates exactly from public data, which is more than can be said for many published composites. Independent auditing of this index is cheap, and the Prosperity Institute deserves credit for that. The point ranks in the middle of the table, however, overstate the precision of the index. A published rank near the middle of the table is better represented by a band of roughly ±15 to 20 places. Small differences in the ranks of countries in the middle of the table are noise, and any statement built on such a difference should be checked against the grid-prior perturbation before it is repeated. The uncertainty in the index localizes exactly where the index is contestable. The near-redundant economic and institutional pillars have wide leeway in their weights. What the ranking actually depends on is how much voice the weakly correlated pillars of Personal Freedom, Natural Environment, and Social Capital are given. Equal weighting resolves this question silently in favor of these pillars receiving full voice, and neither the designers of the index nor the users of its ranking typically recognize that such a decision has been made.

The audit itself takes only a few minutes to complete on a laptop, so there is no computational excuse for publishing point ranks alone. Rank-uncertainty intervals generated as in Section 4.1 would give readers a quick look at which comparisons the index can support, and a one-page table listing the sensitivity of the headline ranks to the alternative choices of normalization and aggregation would give readers the same information for those construction choices. The audit itself is easily transferred to any composite index whose top-level aggregation is published, and the discipline it enforces, exact replication before perturbation, is worth making a standard part of the publication of any such index.

---

7. Limitations

Only the pillar-to-index step is perturbed. The selection of the underlying indicators, the imputation of missing values, the distance-to-frontier normalization of the raw indicators, and all aggregation below the pillar level are held at their published values, so that the rank intervals reported here are lower bounds on the total structural uncertainty of the index. The uniform prior is deliberately diffuse; readers who find it too permissive should anchor on the grid-prior numbers ($\bar{R}_S = 2.93$), which more closely reflect how index designers actually adjust weights in practice. 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 audit is a single cross-section of the 2023 edition of the index. Performing the same audit on each edition would reveal the stability of the sensitivity structure over time.

---

8. Conclusion

The ranking is globally stable but locally soft. The median 90-percent rank interval is 22 places under a fully agnostic prior on the weights and reaches 55 places in the middle of the table. Rank variance concentrates in the weights on the Personal Freedom and Natural Environment pillars and in the aggregation rule, and the leverage of a weight is governed by the independence of its pillar from the rest of the composite index (Spearman -0.76 between S_T and $\hat{\eta}^2$). The equal weighting of the pillars is a substantive decision about how much that independent information counts. Composite rankings should be published together with the uncertainty statements that make such decisions visible to the reader.

Conflict of Interest

The authors declare no conflicts of interest.

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 $X = (X_1, \ldots, X_{k'})$, the variance decomposes as $\text{Var}(Y) = \sum_j V_j + \sum_{j<l} V_{jl} + \cdots$ with $V_j = \text{Var}[E(Y|X_j)]$ (Sobol', 2001). The first-order and total-effect indices are

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

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

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

Correlation ratio (Section 4.3). For pillar j with score column s_j, cut into $B = 10$ quantile bins $G_1, \ldots, G_B$, 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 $\text{Dir}(\mathbf{1}_k)$ has density $\Gamma(k)$, 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 $U_j \overset{\text{i.i.d.}}{\sim} \text{Uniform}(0,1)$ then $-\ln U_j \overset{\text{i.i.d.}}{\sim} \text{Exponential}(1)$, and the normalized vector of k independent Exponential(1) draws is exactly Dir$(\mathbf{1}_k)$. 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 $w \sim \text{Dir}(c \cdot \mathbf{1}_k)$, $E[w_j] = 1/k$ and Var$(w_j) = (k-1)/(k^2(ck+1))$. With $c = 50$ and $k = 12$, sd$(w_j) \approx 0.011$ around $1/12 \approx 0.083$.

Appendix B. Normalization and Aggregation Alternatives

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

- **Identity.** $N_1(s_{ij}) = s_{ij}$, the native distance-to-frontier scale.
- **Min–max.** $N_2(s_{ij}) = 100(s_{ij} - s_j^{\min})/(s_j^{\max} - s_j^{\min})$, 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.** $A_1(m_i; w) = \sum_j w_j m_{ij}$: fully compensatory, a deficit on one pillar offsets one-for-one against a surplus on another.
- **Weighted geometric mean.** $A_2(m_i; w) = \exp(\sum_j w_j \ln(m_{ij} + \delta_j))$: 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

Input	S_i	S_{Ti}
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 $1/12 \approx 0.083$ throughout.

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