

The Additive Audit: Systematic Measurement Error from Additive Aggregation of Non-Substitutable Dimensions

David Kirsch

July 2026

Abstract

When the dimensions of a composite index are non-substitutable—meaning that no amount of improvement on one dimension can compensate for collapse on another—the uniqueness theorem of multiplicative composition (Kirsch 2026) proves that the only valid aggregation function is multiplicative. Any additive aggregation of such dimensions commits a provable composition error that systematically overstates quality by hiding zeros inside averages. We formalize a replicable audit protocol and apply it to two of the world’s most influential composite indices: the V-Dem Electoral Democracy Index and the UNDP Human Development Index. The V-Dem audit reveals 2,586 country-year observations where the additive polyarchy index scores above 0.5 (“democratic”) while the multiplicative index scores below 0.1 (“zero on at least one dimension”)—a 9.6% hidden-zero rate across 26,954 observations. The maximum gap is 0.754 (Hong Kong, 1995: additive score 0.754, multiplicative score 0.000), with the collapsed dimension invariably being elected officials (*v2x_elecoff* = 0). The gap has widened from a mean of 0.23 in the 1900s to 0.35 in the 2010s, as more countries acquire some democratic institutions while retaining authoritarian control of executive selection. The HDI audit confirms the direction of the error (additive > geometric for all countries in all years) but shows smaller magnitude (mean gap 0.007), consistent with the fact that the UNDP’s dimension indices are constructed to avoid zeros. These results provide the first systematic, cross-domain, empirically quantified demonstration that additive aggregation of non-substitutable dimensions produces a directionally predictable measurement error—one that a portable diagnostic protocol can detect in any composite index.

1 Introduction

Composite indices pervade empirical social science, policy evaluation, and institutional decision-making. The Human Development Index ranks 206 countries. The V-Dem Electoral Democracy Index scores 180 countries across 224 years. ESG composite scores guide trillions of dollars in investment. University rankings determine institutional prestige. Credit scores adjudicate billions of lending decisions. Each of these indices aggregates multiple dimensions into a single score.

The aggregation function matters. When dimensions are substitutable—when improvement on one can genuinely compensate for decline on another—additive aggregation (arithmetic mean, weighted sum) is defensible. But when dimensions are non-substitutable—when all are necessary conditions for the composite property—the uniqueness theorem of multiplicative composition proves that the only aggregation function satisfying zero-collapse, continuity, monotonicity, homogeneity, and separability is the multiplicative form $f = k \prod x_i^{\alpha_i}$ (Kirsch, 2026).

The practical consequence of applying additive aggregation to non-substitutable dimensions is that a zero on any dimension is hidden inside an average rather than annihilating the composite score. A country with zero press freedom but high scores on all other democratic dimensions receives a composite democracy score above 0.5 under additive aggregation—the index calls it “moderately democratic” when a single collapsed dimension means it is not democratic at all.

This paper turns the uniqueness result into a portable diagnostic tool. We formalize an audit protocol (Section 2), apply it to the V-Dem Electoral Democracy Index (Section 3) and the Human Development Index (Section 4), and synthesize the findings (Section 5). Our contribution is not the critique of any individual index—prior work has raised the additive-vs-multiplicative question for specific indices (Sagar & Najam, 1998; James, 2008; Munda, 2013). Our contribution is the systematic, cross-domain application of a protocol grounded in a uniqueness proof, with quantified measurement error.

1.1 Prior Art and the Gap This Paper Fills

The question of additive versus multiplicative aggregation has a long history in composite indicator methodology. Sagar & Najam (1998) argued that the HDI’s arithmetic mean “runs counter to the notion of [dimensions] being essential and therefore non-substitutable.” James (2008) proposed multiplicative aggregation for digital preparedness indices. Munda (2013) surveyed compensatory versus non-compensatory aggregation methods for well-being measurement. Paruolo et al. (2012) demonstrated that nominal weights in composite indicators

rarely match their actual importance. Bjerre et al. (2019) showed aggregation sensitivity in immigration policy indices. The OECD’s *Handbook on Constructing Composite Indicators* (OECD, 2008) acknowledges the issue but offers no formal criterion for choosing between aggregation methods.

These are scattered, index-by-index critiques without: (1) a mathematical proof that multiplicative is the *unique* correct form for non-substitutable dimensions; (2) a systematic cross-domain audit applying the same diagnostic to multiple indices; (3) quantified measurement error—how many entities are misclassified; or (4) a connection to the O-Ring inversion result showing what the composition error implies for resource allocation. This paper addresses gaps (1)–(3), building on the uniqueness proof established in Kirsch (2026).

2 The Audit Protocol

Definition 1 (Additive Audit Protocol). *For any composite index C aggregating dimensions $x_1, \dots, x_n$:*

1. **Substitutability assessment.** *Determine whether the dimensions are substitutable. If a zero on dimension x_i should logically annihilate the composite property—if no amount of x_j can compensate—the dimensions are non-substitutable.*
2. **Dual computation.** *Compute both the additive aggregation $C_A = \frac{1}{n} \sum_{i=1}^n x_i$ and the multiplicative aggregation $C_M = (\prod_{i=1}^n x_i)^{1/n}$ for all entities.*
3. **Divergence identification.** *Identify the divergence zone: entities where C_A and C_M disagree on classification. The critical cases are hidden zeros: entities where C_A exceeds a threshold τ but $C_M < \epsilon$, indicating a collapsed dimension masked by the additive average.*
4. **Error quantification.** *Count the number of misclassified entities, compute the mean and distribution of the gap $C_A - C_M$, and identify the binding dimension (the near-zero component driving the divergence).*

The gap $C_A - C_M \geq 0$ always, with equality if and only if all dimensions are equal (by the AM-GM inequality). The gap is largest when dimensions are most unequal—precisely the cases where a binding constraint exists and additive aggregation is most misleading.

3 Audit 1: V-Dem Electoral Democracy Index

3.1 Background

The Varieties of Democracy (V-Dem) project provides the world’s most comprehensive dataset on democracy, covering 180 countries from 1789 to the present with over 450 indicators coded by more than 3,700 country experts (Coppedge et al., 2018). The Electoral Democracy Index (EDI, *v2x_polyarchy*) is V-Dem’s core measure, aggregating five component indices:

- *v2x_elecoff*: Elected officials (are the chief executive and legislature selected through elections?)
- *v2xel_frefair*: Clean elections (free and fair electoral process)
- *v2x_suffr*: Share of population with suffrage
- *v2x_freexp_altinf*: Freedom of expression and alternative sources of information
- *v2x_frassoc_thick*: Freedom of association (thick)

V-Dem computes both a multiplicative polyarchy index (*v2x_mpi*) and an additive polyarchy index (*v2x_api*), then averages them equally to produce the official EDI:

$$EDI = 0.5 \times v2x_mpi + 0.5 \times v2x_api \quad (1)$$

The V-Dem codebook explicitly acknowledges the tension: “We have no strong reason to prefer the additive terms to the multiplicative term, so we give them equal weight” (Coppedge et al., 2018). They further note that “even when the [multiplicative index] is zero, the additive index can achieve as high a score as 0.77.”

This makes V-Dem the ideal test case for the additive audit: V-Dem itself computes both aggregations, acknowledges the divergence, and provides all component data.

3.2 Substitutability Assessment

Are the five dimensions of electoral democracy substitutable? Can a country with zero elected officials (*v2x_elecoff* = 0) compensate with high scores on freedom of expression and clean elections? The answer is clearly no: if the chief executive is not elected, the polity is not an electoral democracy regardless of how free its press is or how fairly its (irrelevant) elections are conducted. The dimensions are non-substitutable. By the uniqueness theorem, the multiplicative aggregation is the correct form.

3.3 Results

We analyzed V-Dem v16 Core data: 26,954 valid country-year observations with complete data on both $v2x_api$ and $v2x_mpi$.

3.3.1 Hidden Zeros

Of 26,954 country-year observations, **2,586 (9.6%)** have $v2x_api > 0.5$ but $v2x_mpi < 0.1$. These are countries that the additive index classifies as “moderately democratic” but that have a collapsed dimension—the multiplicative index identifies them as having a zero that the additive score hides.

3.3.2 Gap Statistics

Statistic	Value
Total observations	26,954
Mean gap ($C_A - C_M$)	0.269
Median gap	0.257
Maximum gap	0.754
Hidden zeros ($v2x_api > 0.5$, $v2x_mpi < 0.1$)	2,586 (9.6%)
Observations with gap > 0.1	76.1%
Observations with gap > 0.3	42.9%

Table 1: V-Dem additive–multiplicative gap statistics, all country-years.

The mean gap of 0.269 means that, on average, the additive index overstates democracy by more than a quarter of the 0–1 scale. Over three-quarters of all observations have a gap exceeding 0.1.

3.3.3 Extreme Cases

Country	Year	$v2x_api$	$v2x_mpi$	Gap	Binding Dimension
Hong Kong	1995	0.754	0.000	0.754	Elected officials (0)
Hong Kong	1996	0.754	0.000	0.754	Elected officials (0)
Japan	1951	0.751	0.000	0.751	Elected officials (0)
Japan	1950	0.750	0.000	0.750	Elected officials (0)
Latvia	1920	0.749	0.000	0.749	Elected officials (0)
Latvia	1921	0.748	0.000	0.748	Elected officials (0)

Table 2: Top divergence cases in V-Dem. All have $v2x_elecoff = 0$ (chief executive not elected) but high scores on all other dimensions.

The diagnostic case is Hong Kong (1990s): freedom of expression scored 0.897, freedom of association 0.847, universal suffrage 1.0, and clean elections 0.798—but the chief executive was not elected ($v2x_elecoff = 0$). The additive index produces 0.754 (“substantially democratic”), while the multiplicative index correctly produces 0.000 (“not an electoral democracy”). V-Dem’s official EDI, averaging both, reports 0.377—a score that obscures the absolute zero on the most fundamental dimension.

3.3.4 Temporal Trend

Decade	Observations	Mean Gap	Max Gap	Extreme Count (>0.3)
1900s	1,255	0.231	0.741	427
1950s	1,559	0.276	0.751	646
1970s	1,574	0.283	0.735	698
1990s	1,746	0.344	0.754	1,065
2000s	1,773	0.351	0.743	1,095
2010s	1,789	0.357	0.739	1,138
2020s	1,074	0.354	0.577	693

Table 3: V-Dem gap by decade. The divergence has widened over time.

The mean gap has increased from 0.231 in the 1900s to 0.357 in the 2010s. This trend reflects the expansion of partial democratization: more countries have acquired some democratic institutions (freedom of expression, civil society, elections) while retaining authoritarian control of executive selection. The additive index registers this as democratic progress; the multiplicative index recognizes that a zero on the fundamental dimension nullifies the gains.

3.3.5 Predictive Validation: Do Hidden Zeros Predict Autocratic Outcomes?

The descriptive audit shows that 2,586 country-years are misclassified by the additive index. But is the multiplicative index actually the better predictor? We test this by tracking regime outcomes. For each country-year, we use V-Dem’s Regimes of the World classification ($v2x_regime$: 0=closed autocracy, 1=electoral autocracy, 2=electoral democracy, 3=liberal democracy) and compare outcomes at 5-year and 10-year horizons for three groups:

- **Hidden zeros** ($v2x_api > 0.5$, $v2x_mpi < 0.1$): additive says democratic, multiplicative says zero.
- **Agreement-democratic** ($v2x_api > 0.5$, $v2x_mpi > 0.5$): both indices agree the country is democratic.

Group	5-Year Horizon		10-Year Horizon	
	n	Outcome	n	Outcome
Hidden zeros	1,674	91.1% autocratic	1,625	83.6% autocratic
Agreement-dem	3,119	1.6% breakdown	2,820	2.7% breakdown

Table 4: Regime outcomes by index agreement group. Hidden-zero countries—classified as “democratic” by the additive index but as zero by the multiplicative index—remain autocratic 91.1% of the time at 5 years. The multiplicative index is correct; the additive index is systematically wrong.

The result is unambiguous: countries that the additive index calls “moderately democratic” (score > 0.5) but that the multiplicative index identifies as having a collapsed dimension (score < 0.1) remain autocratic **91.1%** of the time within 5 years and **83.6%** within 10 years. By contrast, countries where both indices agree on democracy experience breakdown only 1.6% of the time at 5 years (2.7% at 10 years).

The additive index misclassifies these countries as democratic; subsequent outcomes confirm they are not. The multiplicative index, by collapsing the composite score to zero when any dimension is zero, correctly predicts their autocratic character.

Case study: Hong Kong. Hong Kong illustrates the diagnostic power of the multiplicative index. From 1990 to 2020, its additive polyarchy score ranged from 0.509 to 0.754—the additive index consistently called it “moderately to substantially democratic.” Its multiplicative score was 0.000 for the entire period (elected officials = 0: the chief executive was never elected). V-Dem’s own regime classification confirms the multiplicative index: Hong Kong was coded as a *closed autocracy* ($v2x_regime = 0$) for every year from 1990 to 2020. The additive index was wrong for three decades running. Moreover, even the non-zero dimensions eventually collapsed: freedom of expression fell from 0.897 in 1995 to 0.428 in 2020, as Beijing tightened control. The multiplicative index identified Hong Kong’s democratic deficit from the start; the additive index masked it until the other dimensions collapsed too.

4 Audit 2: Human Development Index

4.1 Background

The UNDP’s Human Development Index aggregates three dimensions—health (life expectancy), education (expected and mean years of schooling), and income (GNI per capita, log-transformed)—into a single measure of human development for 206 countries since 1990. In 2010, the UNDP

switched from arithmetic to geometric mean aggregation, implicitly acknowledging the substitutability critique raised by Sagar & Najam (1998) and others—though without citing a formal justification.

This switch creates a natural experiment: we can compare pre-2010 additive HDI with geometric recomputation to quantify the effect of the aggregation change.

4.2 Substitutability Assessment

Are health, education, and income substitutable for human development? Can infinite income compensate for zero education? In the extreme, clearly not—the dimensions are non-substitutable. However, the UNDP constructs its dimension indices with fixed bounds that prevent true zeros: life expectancy is mapped to $[0, 1]$ via $LE_{idx} = (LE - 20)/(85 - 20)$, education uses $(EYS/18 \times MYS/15)^{1/2}$, and income uses a log transformation. These bounded transformations compress the range and prevent the extreme cases that drive the largest divergences.

4.3 Results

We analyzed 206 countries across 7 years spanning the pre-2010 (additive-official) and post-2010 (geometric-official) periods: 2000, 2005, 2009, 2010, 2015, 2020, 2022.

Year	Countries	Mean Gap	Max Gap	Max Gap Country
2000	187	0.0103	0.0577	Yemen
2005	199	0.0086	0.0498	Senegal
2009	200	0.0075	0.0456	Niger
2010	202	0.0073	0.0456	Niger
2015	203	0.0064	0.0473	Niger
2020	203	0.0055	0.0445	Niger
2022	204	0.0057	0.0460	Niger

Table 5: HDI additive–geometric gap. The gap is small (0.005–0.010) because UNDP dimension indices are bounded away from zero.

The Spearman rank correlation between additive and geometric rankings exceeds 0.998 for all years, confirming that the HDI’s bounded construction makes the two aggregations nearly equivalent. No country crosses the extreme divergence threshold ($C_A > 0.5$ and $C_M < 0.4$).

However, the gap is systematic: it is always positive (additive $>$ geometric), always largest for countries with the most unequal dimension scores, and always driven by the **education**

dimension. Niger, Burkina Faso, Mali, Senegal, and Yemen consistently occupy the top divergence positions, all with education as the binding dimension. Education is the binding dimension in 100% of top-10 divergence cases across all seven years analyzed.

4.4 Interpretation

The HDI is a *weak* test case for the additive audit, and this is itself informative. The UNDP’s dimension construction prevents true zeros, compressing all countries into a region where the AM-GM inequality produces minimal divergence. The 2010 switch to geometric mean was the correct direction—our analysis confirms the systematic (if small) upward bias of the arithmetic mean—but the switch addresses the problem only at the top-level aggregation. Within the education dimension, expected and mean years of schooling are combined geometrically (post-2010), but a country with zero years of mean schooling would still have its education index pulled toward zero rather than annihilated, because the component is bounded.

The comparison between V-Dem (large divergence, true zeros possible) and HDI (small divergence, zeros prevented by construction) illustrates a general principle: *the magnitude of additive composition error depends on whether the index permits zeros*. Indices whose dimensions are bounded away from zero will show the error only in attenuated form. Indices whose dimensions can reach zero—like V-Dem’s elected officials component—expose the full force of the composition error.

5 Cross-Audit Synthesis

5.1 Systematic Direction of Error

Both audits confirm that additive aggregation produces a *directionally predictable* error: the additive score always exceeds the multiplicative/geometric score ($C_A - C_M \geq 0$, by AM-GM), with the gap largest when dimensions are most unequal. This is not a random measurement error that averages out; it is a systematic bias that always overstates the composite score relative to the binding constraint.

5.2 When Does the Error Matter?

The error matters when:

1. Dimensions can reach zero (V-Dem: 9.6% hidden-zero rate; HDI: prevented by construction)

2. Classification decisions depend on threshold scores (a country classified as “democratic” or “high human development” may be misclassified if its score is inflated by additive averaging of non-substitutable dimensions)
3. Resource allocation follows from the score (the O-Ring inversion result from Kirsch (2026) shows that the binding constraint—the lowest dimension—is where marginal investment produces the highest return; additive aggregation obscures exactly this dimension)

5.3 The Audit as a Portable Diagnostic

The protocol in Section 2 can be applied to any composite index. The key diagnostic questions are:

1. Are the dimensions substitutable?
2. Does the index permit zeros?
3. What is the gap distribution?

If the answer to (1) is no and (2) is yes, the index is systematically inflated by additive aggregation, and the magnitude is given by (3). Candidate indices for future audits include FICO credit scores, Basel systemic risk indicators, ESG composites, Pugh matrices used in engineering design evaluation, and university league tables.

6 Discussion

The V-Dem result is particularly striking because V-Dem’s own codebook documents the divergence. Coppedge et al. (2018) note that “even when the [multiplicative index] is zero, the additive index can achieve as high a score as 0.77.” Our audit confirms this with comprehensive data: the maximum observed gap is 0.754, and 2,586 country-years are classified as “moderately democratic” by the additive index when the multiplicative index identifies a completely collapsed dimension.

V-Dem’s decision to average the two indices equally—“we have no strong reason to prefer the additive terms to the multiplicative term”—is understandable in the absence of a formal criterion. The uniqueness theorem provides that criterion: when dimensions are non-substitutable (and elected officials are plainly non-substitutable with freedom of expression for electoral democracy), the multiplicative form is not a preference but a mathematical necessity.

The UNDP’s 2010 switch from arithmetic to geometric mean was directionally correct and is sometimes cited as evidence that the aggregation question is settled. Our analysis shows it is not: the HDI’s bounded dimension construction prevents the large divergences that the V-Dem analysis reveals. The question is not whether to use geometric mean (the UNDP answered this in 2010) but whether dimension indices should be constructed to prevent zeros in the first place—and if so, whether this prevention obscures genuine binding constraints.

The temporal trend in the V-Dem data—mean gap rising from 0.23 to 0.35 over a century—has substantive implications for the study of democratization. The increasing gap reflects what Levitsky and Way (2010) call “competitive authoritarianism”: regimes that maintain many democratic institutions while controlling the most fundamental one (executive selection). Additive indices register competitive authoritarian regimes as partially democratic; multiplicative indices recognize that the controlled dimension nullifies the democratic character of the others.

7 Conclusion

We have demonstrated that additive aggregation of non-substitutable dimensions produces a systematic, directionally predictable measurement error. The audit protocol formalized here can be applied to any composite index. Applied to V-Dem, it reveals 2,586 country-years of hidden zeros (9.6% of all observations) where the additive index calls a country “democratic” despite a collapsed dimension. Applied to the HDI, it confirms the direction of the error but shows smaller magnitude due to the prevention of true zeros by construction.

The practical implication is a concrete recommendation: any organization constructing or using a composite index of non-substitutable dimensions should compute both additive and multiplicative scores, report the gap, and use the multiplicative score for decisions that depend on all dimensions being present. This is not a methodological preference; it follows from the uniqueness theorem. Every institution currently using additive composition of non-substitutable dimensions is making a systematic error that always biases in the same direction: hiding zeros inside averages.

References

- Bjerre, L., Römer, M., & Zobel, M. (2019). Aggregation sensitivity in immigration policy indices. *Comparative European Politics*, 17(4), 587–608.

- Coppedge, M., Gerring, J., Knutsen, C.H., Lindberg, S.I., Teorell, J., et al. (2018). V-Dem Methodology v8. *V-Dem Working Paper Series*, 2018:57.
- James, J. (2008). Digital preparedness versus the digital divide: A proposal for a new composite index. *Technology Analysis & Strategic Management*, 20(1), 79–95.
- Kirsch, D. (2026). Multiplicative composition of independent dimensions in emergent systems. *Working paper*.
- Munda, G. (2013). Beyond GDP: An overview of measurement of progress with a focus on well-being indicators. *Scientometrics*, 94(3), 1147–1165.
- OECD (2008). *Handbook on Constructing Composite Indicators: Methodology and User Guide*. Paris: OECD Publishing.
- Paruolo, P., Saisana, M., & Saltelli, A. (2012). Ratings and rankings: voodoo or science? *Journal of the Royal Statistical Society: Series A*, 176(3), 609–634.
- Sagar, A.D. & Najam, A. (1998). The human development index: A critical review. *Ecological Economics*, 25(3), 249–264.
- Stiglitz, J.E., Sen, A., & Fitoussi, J.P. (2009). *Report by the Commission on the Measurement of Economic Performance and Social Progress*. Paris.