Does life-satisfaction inequality measure societal inequality? A focal-value-rounding critique

C.P. Barrington-Leigh*

First draft: Solstice 2026.

Compiled July 9, 2026

Please use the latest version of this paper online.

Abstract

The within-country dispersion of self-reported life satisfaction has been proposed and used as a comprehensive measure of societal inequality. A negative cross-country association between mean life satisfaction and its standard deviation has been read as evidence that this inequality is itself welfare-relevant. Existing measurement critiques of life satisfaction include the nonlinearity and boundedness of response scales. I describe a further, distinct problem: a substantial and predictable share of respondents simplify the 0–10 scale to the focal subset $\{0, 5, 10\}$ — "focal-value rounding" (FVR) — a behaviour that breaks not only cardinality but ordinality of the reported distribution, and concentrates reported mass at those three values. I derive a sharp bound on the standard-deviation bias that a focal-responding population can induce; at empirically typical FVR fractions the admissible bias is roughly half the cross-country range of measured dispersion. I then estimate and correct for the realized bias by fitting a model of FVR behavior — an adaptation of the model in Barrington-Leigh (2024) — to the Gallup World Poll Cantril ladder (135 countries; $N = 300{,}137$) as well as to three other multi-country life evaluations (a second Gallup item and two Global Flourishing Study items). The estimated corrections are material in level: FVR inflates measured standard deviation in 133 of 135 countries, by a median of 0.090 points. However, because the corrections are nearly uniform in sign across countries, rankings of the standard deviation survive essentially intact. The ordinal inequality indices which Cowell and Flachaire (2017) advocate as the theoretically appropriate alternative are the statistics most disturbed by FVR correction. The Goff et al. (2018) mean–dispersion correlation survives the FVR correction with modest attenuation. Thus FVR does not, by itself, overturn previously reported correlations; it does, however, add a further reason to remain wary of — or to avoid — scalar measures of subjective wellbeing inequality: the candidate statistics are differently contaminated, the ordinal repairs no less than the cardinal originals, and they disagree with one another. Lastly, I show that FVR-correction can improve the pairwise dominance-comparability of full distributions.

*Author contact at https://wellbeing.research.mcgill.ca/contact. Comments welcome.
Interactive demonstrations available online. This work was supported by SSHRC grant 435-2024-1219. I am grateful to Gaston Yalonetzky and audiences in Vancouver and Luxembourg for helpful comments.

Contents

List of Figures

List of Tables

1 Introduction

If self-reported life satisfaction is accepted as a comprehensive welfare measure — one that captures health, relationships, security, and purpose, and other human experience in appropriate measure along with income — then its distribution within a population is of as much interest as its mean. Accordingly, researchers have computed inequality statistics directly from the distribution of life satisfaction responses: Goff et al. (2018, hereafter, GHM) use the within-country standard deviation; Stevenson and Wolfers (2008) use the Gini coefficient; practitioner guidance recommends a portfolio including the standard deviation, threshold shares, the Gini, and the Shannon index (What Works Centre for Wellbeing, 2018); and a methodological literature has debated which statistic is proper to the task at least since Kalmijn and Veenhoven (2005), with Bayesian treatments of the cross-national response distribution appearing in Hasegawa and Ueda (2011). The choice among distribution-sensitive aggregates inevitably encodes an ethical stance, and Panasiuk et al. (2025) show that life satisfaction responses depart from normality in every one of 148 countries — so that no mean-plus-dispersion summary captures the distribution.

The stakes have been raised by a substantive empirical claim: Goff et al. (2018) report that country-mean life satisfaction is lower where the standard deviation of life satisfaction is higher, conditional on income, and read this as evidence that wellbeing inequality is itself welfare-reducing, and more so than income inequality. Two claims are therefore under test when a dispersion statistic of life satisfaction is reported: that cross-country variation in the statistic meaningfully reflects variation in some underlying wellbeing inequality; and that this inequality is welfare-relevant in the way the mean–dispersion association suggests.

Doubt about the first claim is natural given the analogous concern about cardinality of the mean, though the latter issue turns out to have limited empirical impact (Ferrer-i-Carbonell and Frijters, 2004; Kaiser and Vendrik, 2020). An ordinal-data literature has produced, most recently, the status-based index family of Cowell and Flachaire (2017) and its associated dominance theory (Jenkins, 2021; Gravel et al., 2021). Grimes et al. (2023) bring this machinery to the GHM question and find that the negative mean–dispersion association is not robust: it vanishes when country and wave fixed effects are included, and the theoretically appropriate ordinal indices yield relationships of opposite signs depending on whether status is computed looking upward or downward.

This paper describes a further measurement problem, distinct from ordinality, and quantifies its consequences for both claims. Barrington-Leigh (2024) documents that a substantial share of respondents simplify a 0–10 response scale to the focal subset $\{0, 5, 10\}$, and that the propensity to do so is predictable — most strongly by education. This "focal-value rounding" (FVR, a form of response heaping in subjective numeric data) breaks not just the cardinality of the reported scale but its ordinality: a respondent who maps an interior assessment to 5 produces a report that no monotone transformation of the labels can undo. FVR of this kind is plainly visible in published cross-country response histograms (e.g., Figure 4 of Grimes et al., 2023), though never modeled in the literature on life satisfaction inequality.

This paper makes four contributions. First, theory: I derive a bound on the bias that FVR can induce in the standard deviation of reported life satisfaction — the dispersion analogue of the mean-bias bound in Barrington-Leigh (2024, Proposition 1) — and show in simulation that the bias can take either sign, that it varies with the latent mean, and that it propagates undiminished into the Cowell–Flachaire ordinal indices. The reason is that those indices are functionals of the response CDF, which is necessarily changed by FVR; an index immune to scale-stretching is not FVR-proof. In a synthetic cross-section of countries with no genuine wellbeing-inequality channel (latent dispersion identical across countries by construction), heterogeneous FVR rates nonetheless contaminate the cross-country mean–dispersion comparison, spuriously reproduce GHM's inequality-aversion interaction for the upper range of the life satisfaction scale, and contaminate the ordinal variation-ratio measure.

Second, measurement: I adapt the model of Barrington-Leigh (2024) to allow for separate, predictable rounding propensities toward each focal value, layered on an ordered probit with a hierarchical per-country latent scale. I fit this to four large multi-country datasets: the Gallup World Poll Cantril ladder (2020–2022; 135 countries; $N = 300{,}137$), the Gallup satisfaction-with-life item (2007–2010; 114 countries; $N = 109{,}375$), and the Global Flourishing Study life-today and satisfaction items (2023; 22 countries; $N = 177{,}353$ and 177,157). Each fit yields, per country, the latent (FVR-free) response distribution alongside the model's predicted reported distribution, with full posterior uncertainty.

Third, empirics: for five inequality statistics I compare cross-country values and rankings before and after the FVR correction. I re-run the GHM country-level regression with raw, predicted, and corrected dispersion regressors; and I report dominance comparisons (Jenkins, 2021) on the corrected distributions.

Fourth, an argument about measures: I argue on a priori grounds (Section 2.2) that the Gini coefficient, the coefficient of variation, and the share reporting below 5 — all in use in this literature — are unsuitable for a bounded scale with arbitrary origin: the first two are mechanically linked to the mean by construction, and the third saturates toward its bounds as the mean moves away from its fixed threshold; I accordingly exclude all three from the main analysis of the mean–dispersion relationship.

The empirical findings are easily summarized, and are on balance more benign than the theory might suggest. The corrections are real and material in level: for instance, FVR inflates measured standard deviation in 133 of 135 Gallup-ladder countries. But the corrections are nearly uniform in sign, and much of their variation is shared across countries, so country rankings based on the standard deviation are largely unaffected, and the GHM coefficient survives with modest attenuation. The Cowell–Flachaire ordinal indices — the theoretically preferred instruments — are the most disturbed. FVR alone is thus not sufficient to overturn the existing cross-sectional correlations; yet it does further undermine the case for reading any scalar of the reported distribution as a measurement of wellbeing inequality, since the candidate statistics disagree with each other and are differently contaminated, and those with the most theoretical appeal are contaminated the most.

The rest of the paper proceeds as follows. Section 2 reviews the statistics in use and argues for excluding the mean-linked ones. Section 3 presents the measurement model. Section 4 contains the bound, the analytic bias curves, and a synthetic falsification of GHM's inequality-aversion test. Section 5 presents the global estimates, the corrected rankings, the GHM re-estimation, and the dominance analysis. Section 6 weighs what survives, and Section 7 concludes.

2 Statistics of life-satisfaction inequality

2.1 The measures in use

Life satisfaction surveys produce, for each country, an empirical distribution $\{p_k\}_{k=0}^{10}$ over an eleven-point scale.[1] I focus on five statistics of that distribution. The within-country **standard deviation** is the measure on which the headline mean–dispersion claims rest (Goff et al., 2018). The **80:20 difference** is the difference between the survey-weighted mean response of the top and the bottom quintile of a response distribution, and is among the measures recommended in practitioner guidance (What Works Centre for Wellbeing, 2018). The status-based indices I_0^D **and** I_0^U of Cowell and Flachaire (2017) arose from an ordinal-index lineage Allison and Foster (2004), Abul Naga and Yalcin (2008), and Apouey (2007); in these, each respondent is scored by her upward- or downward-looking *status* — e.g., the share of the population responding no higher than she does — rather than by the numeric label. The **variation ratio** is $1 - \max_k p_k$, i.e., the share of responses away from the modal category. This is the purely ordinal measure that GHM themselves deploy in defence of their result (Goff et al., 2018). Each of the five is computed below on synthetic and on empirical data and compared with its FVR-corrected value. Dominance criteria for ordinal distributions, applied in Section 5, are supplied by Jenkins (2021) and Gravel et al. (2021).

Other statistics have been applied elsewhere in the context of life satisfaction inequality: for instance, Stevenson and Wolfers (2008) use the Gini coefficient; What Works Centre for Wellbeing (2018) suggests the coefficient of variation, threshold shares, and the Shannon index; and Delhey and Kohler (2011) proposes the instrument-corrected standard deviation. Three of these are excluded from the main analysis for the reasons developed next.

2.2 Why the Gini, CV, and threshold share are excluded

The Gini coefficient and the coefficient of variation are *relative* inequality measures: both are ratios with the mean in the denominator, and both are therefore invariant to a rescaling of all values by a common factor. That invariance is their point. For a ratio-scale quantity such as income — unbounded above, with a meaningful zero — scale invariance is exactly the right requirement: a currency conversion should not change measured inequality.

A 0–10 response scale has neither property. Its origin is arbitrary: the same survey item could have been printed 1–11, and nothing in the responses distinguishes the two labellings. A statistic that is invariant to multiplying all responses by a constant, but *not* to adding one, imports the

[1] 0–10 scales are now standard (OECD, 2025), though the World Values Survey still uses a 1–10 scale, and smaller, Likert-style response scales with verbal descriptions for each response option also exist in some surveys.

geometry of a ratio scale onto data that possess no zero. The practical consequence is mechanical: with the mean in the denominator and the support bounded, cross-country variation in the mean produces cross-country variation in the Gini and the CV even when absolute dispersion is identical, and the induced association with the mean is negative by construction. A regression of mean life satisfaction on the Gini or CV of life satisfaction therefore has its sign partly built in before any behaviour enters. The standard deviation is not immune to boundedness either — its maximum on a bounded support is itself a function of the mean, the observation behind the corrected statistic of Delhey and Kohler (2011) and the class of normalized indices axiomatized by Permanyer et al. (2025) — but the SD at least does not divide by the mean, and it is the statistic on which the headline empirical claims rest.

A third statistic, the share reporting below 5, is hardly even a measure of inequality when used across populations with varying means. It is the response CDF evaluated at a single, fixed point: as a country's mean moves away from that threshold in either direction, the share below it necessarily approaches 0 or 1, whatever the spread of the remaining mass. This is more than the generic bounded-scale censoring that afflicts every statistic on this scale (Section 4.3): near either tail the statistic is set almost entirely by location, leaving dispersion almost nothing left to determine.

Accordingly, these measures are excluded from the main analysis, although it is reproduced for them in in Appendix E.

3 Focal-value rounding and a model to correct for it

3.1 The phenomenon

Barrington-Leigh (2024) documents, in individual-level data from Canada, the United Kingdom, Australia, and the United States, that responses to 0–10 life satisfaction items concentrate at $\{0, 5, 10\}$; that this FVR reflects a respondent-level propensity to simplify the scale rather than a property of the underlying wellbeing distribution; and that this propensity is strongly predicted by education. Reported national distributions elsewhere show the same spikes (Barrington-Leigh, 2026), including in some of the World Values Survey histograms reproduced by Grimes et al. (2023), and the estimates below imply that the behaviour is ubiquitous in global survey data. A respondent who rounds an interior assessment to a focal value produces a report that no relabelling of the scale can restore: FVR violates ordinality, the assumption that the ordinal-inequality literature retains.

3.2 An adaptation of the Barrington-Leigh (2024) model

The model in Barrington-Leigh (2024) treats each respondent as either using the full scale or restricting to $\{0, 5, 10\}$, with a single estimated propensity.[2] The adaptation used here — which has not been published elsewhere — replaces the single propensity with three: for each focal value $f \in \{0, 5, 10\}$, respondent i retains her full-scale response in the neighbourhood of f with probability

[2] In the World Values Survey, the focal values are {1,5,10}.

$p_i^{H,f}$, and otherwise rounds responses in that neighbourhood to f. I refer to it as the *reallocation* model.

Latent layer. Respondent i in country c has a latent life satisfaction score

$$\eta_{S,i} = \mu_{S,i} + \varepsilon_{S,i}, \qquad \mu_{S,i} \equiv X_{S,i}^\top \beta_{S,c}, \qquad \varepsilon_{S,i} \sim \mathcal{N}\big(0, \sigma_c^2\big), \tag{1}$$

with X_S containing a constant, log household income, an education indicator, and gender. An ordered probit with fixed cardinal cutpoints $c_k = k - 5.5$ for $k = 1, \ldots, 10$, together with $c_0 = -\infty$ and $c_{11} = +\infty$, maps η_S to a full-scale probability mass function (**PMF**) $\boldsymbol{\pi} = (\pi_0, \ldots, \pi_{10})$, with $\pi_k = \Phi\big((c_{k+1} - \mu_{S,i})/\sigma_c\big) - \Phi\big((c_k - \mu_{S,i})/\sigma_c\big)$ for $k = 0, \ldots, 10$ — the Gaussian mass of the latent score falling between consecutive cutpoints. The scale σ_c is the latent dispersion this paper sets out to measure, so, unlike the textbook ordered probit that normalizes it to one, it is kept free and country-specific and enters here standardizing the cutpoint distances. The integer report is thus the interval-censored image of η_S rather than η_S itself.

Rounding layer. With neighbourhoods $\mathcal{N}(0) = \{0, 1, 2\}$, $\mathcal{N}(5) = \{3, \ldots, 7\}$, $\mathcal{N}(10) = \{8, 9, 10\}$ and $p_i^{H,f} = \operatorname{logit}^{-1}(X_{N,i}^\top \beta_{N,f,c})$, the reported PMF is

$$\begin{aligned} P(0) &= \pi_0 + (1 - p^{H,0})(\pi_1 + \pi_2), & P(j) &= p^{H,0}\,\pi_j,\ j \in \{1, 2\}, \\ P(5) &= \pi_5 + (1 - p^{H,5})(\pi_3 + \pi_4 + \pi_6 + \pi_7), & P(j) &= p^{H,5}\,\pi_j,\ j \in \{3, 4, 6, 7\}, \\ P(10) &= \pi_{10} + (1 - p^{H,10})(\pi_8 + \pi_9), & P(j) &= p^{H,10}\,\pi_j,\ j \in \{8, 9\}, \end{aligned} \tag{2}$$

which preserves total mass. X_N contains a constant and an education indicator. Estimation is by Hamiltonian Monte Carlo (NUTS) in PyMC.

Partial-pooling layer. The latent coefficients $\beta_{S,c}$, the focal propensities $\beta_{N,f,c}$, and the latent scale σ_c are country-specific but estimated together through partial pooling. For the question at hand the consequential parameter is the latent scale: $\log \sigma_c$ is drawn from a common hyperdistribution, so each country's latent dispersion is informed by, but not pinned to, the global pattern. In simulation-based calibration, the specification reproduces true cross-country σ_c variation when it is present and finds none when it is absent (Appendix K, Figure K1).

3.3 Identification

In a traditional ordered logit or probit the cutpoints are free parameters, estimated alongside the latent index. This is also true of the Barrington-Leigh (2024) FVR mixture model, and a discussion there treats the issue of FVR identification more generally. In the reallocation model, free cutpoints exacerbate this challenge: a cutpoint bordering a focal value may not be separately identifiable from the rounding propensity toward that value. Specifically, as the top cutpoint shifts outward, the latent mass at 10 shrinks, but a compensating decrease in the constant of β_N raises the rounding propensity, leaving the reported distribution — and hence the likelihood — unchanged along the ridge.

Pinning the cutpoints to the cardinal positions $c_k = k - 5.5$ removes the likelihood ridge. It fixes the latent metric to the units of the response scale (as in OLS), lets σ_c and the focal propensities converge separately, and identifies a constant in X_S. Thus, the FVR correction in this paper is identified under cardinal, equally spaced cutpoints. While addressing the ordinality violation of FVR, the conclusions drawn below about specifically-ordinal indices are conditional on that identifying restriction.

3.4 The objects of analysis

Estimating the model leaves three PMFs over $\{0, \ldots, 10\}$: the raw empirical PMF $\mathbf{P}_c^{\text{obs}}$; the posterior predicted PMF $\hat{\mathbf{P}}_c^{\text{pred}}$, the model's expectation of the reported data and the natural baseline for isolating the correction; and the latent PMF $\hat{\boldsymbol{\pi}}_c$, the counterfactual with $p^{H,f} \equiv 1$, i.e. what would be reported if every respondent used the full scale. For any statistic $T(\cdot)$ I compare $T(\hat{\mathbf{P}}_c^{\text{pred}})$ with $T(\hat{\boldsymbol{\pi}}_c)$ across countries. Because both arise from the same posterior, this comparison isolates the FVR correction, and evaluating both within each posterior draw gives exact credible intervals for any cross-country statistic of the pair.

4 How large can the damage be?

This section asks how large the FVR distortion to a dispersion statistic might be, and it does so without any survey data. I proceed in two steps. First I derive an analytic bound on the standard-deviation bias that holds for any latent distribution. Then, because the realized bias depends on quantities no bound can pin down, I turn to a synthetic simulation of cross-country data — a data-generating process with *no* genuine wellbeing-dispersion channel, specified in full in Appendix C — to show how FVR alone can contaminate cross-country comparisons. Empirical, survey-based estimates follow later, in Section 5.

4.1 A sharp bound on the standard-deviation bias

The following is the dispersion analogue of the mean-bias bound in Barrington-Leigh (2024, Proposition 1). As there, the relevant comparison is between what respondents actually report and what they would report on the full eleven-point scale.

Proposition 1 (Maximum SD bias due to FVR)**.** *Suppose a fraction $\lambda \in [0, 1]$ of respondents report the nearest element of $\{0, 5, 10\}$ (with thresholds at 2.5 and 7.5) instead of the integer they would otherwise give on the full 0–10 scale. Let σ^* be the standard deviation of full-scale responses and σ^{obs} that of the actual, partially rounded responses. Then*

$$|\sigma^{\text{obs}} - \sigma^*| \;\leq\; 2\sqrt{\lambda}, \tag{3}$$

and the bound is sharp for every $\lambda \in [0,1]$ — as an increase or as a decrease, by the λ-indexed distributions given in the proof (Appendix B).

Remark 1. *For* FVR *rates of $\lambda \approx 0.20$–0.40, as estimated in Barrington-Leigh (2024), the bound permits a bias of ∼0.9–1.3 points on the 0–10 scale — roughly half the cross-country range of measured life satisfaction dispersion (Table 2). The two λ-indexed families of distributions that attain the bound (Appendix B) show that* FVR *can either inflate or compress the measured SD, depending on where the population sits relative to the rounding thresholds. The bound is permissive, not predictive; whether the realized bias approaches it, and which sign it takes, are questions for the structural estimates of Section 5.*

4.2 Every dispersion statistic is contaminated by the latent mean

Figure 1 evaluates each dispersion statistic[3] on the exact reported distribution implied by a latent $\mathcal{N}(\mu, \sigma^*)$ — clipped and rounded to the 0–10 integers, then with a fraction λ of mass moved to the nearest focal value — with no sampling noise, and traces it against the mean μ, one curve per focal-responding share λ. The λ=0 curve is the pure bounded-scale (censoring) baseline; its vertical gap to the λ>0 curves is the FVR bias.[4] The left column of the figure fixes the latent width at $\sigma^* = 1$; the right column raises it to 2.35, as motivated by the empirical country data to follow.

At $\sigma^* = 1$, the FVR bias in the standard deviation (Figure 1, top row) is *non-monotone* in the country mean. When the mean sits near the central focal value 5, FVR collapses responses onto a single point, leading to compression of the variance. As the mean rises or falls away from the center to 7.5 or 2.5, the variance is exaggerated by FVR moving responses both to 5 and to an extreme, leading (for $\lambda = 40\%$) to a maximum bias of 0.74. Finally, as the mean approaches either extreme the net effect of FVR is increasingly to collapse the distribution towards that extreme, again diminishing the variance. Because the bias changes sign across the scale, no single correlation coefficient summarizes it and no linear control for the mean can purge it.[5]

The same mean-dependence contaminates the other three statistics, each with its own geometry. The ordinal variation ratio and the two Cowell–Flachaire status indices all vary nonmonotonically with the mean — with peak FVR biases of 0.25, 0.14, and 0.14 respectively. The Cowell–Flachaire indices' invariance to relabelling does not buy invariance to FVR, because FVR is not a relabelling: it reshapes the response distribution.

At $\sigma^* = 1$, the variation ratio furnishes a relatively clean separation of FVR from censoring. That is, between means of 1 and 9, its λ=0 curve is unaffected by censoring. Being a share of responses away from the modal category, it necessarily saw-tooths as the mean crosses each half-integer and the

[3] Two online interactive simulations demonstrate the effects of these measurement problems in shaping SWB reports and in inducing a mechanical relationship between mean and dispersion metrics.

[4] Even this baseline is not exactly the latent SD: discretizing a continuous latent to integers adds Sheppard's correction, so at $\sigma^* = 1$ the uncensored reported standard deviation averages about $1 + \frac{1}{\sqrt{12}} \approx 1.04$ (Sheppard, 1897).

[5] The bias can even be non-monotone in the *degree* of rounding. To see this, consider a latent response that is all "7"s or all "8"s: as the FVR fraction rises from 0 to 1, the standard deviation first rises and then declines.

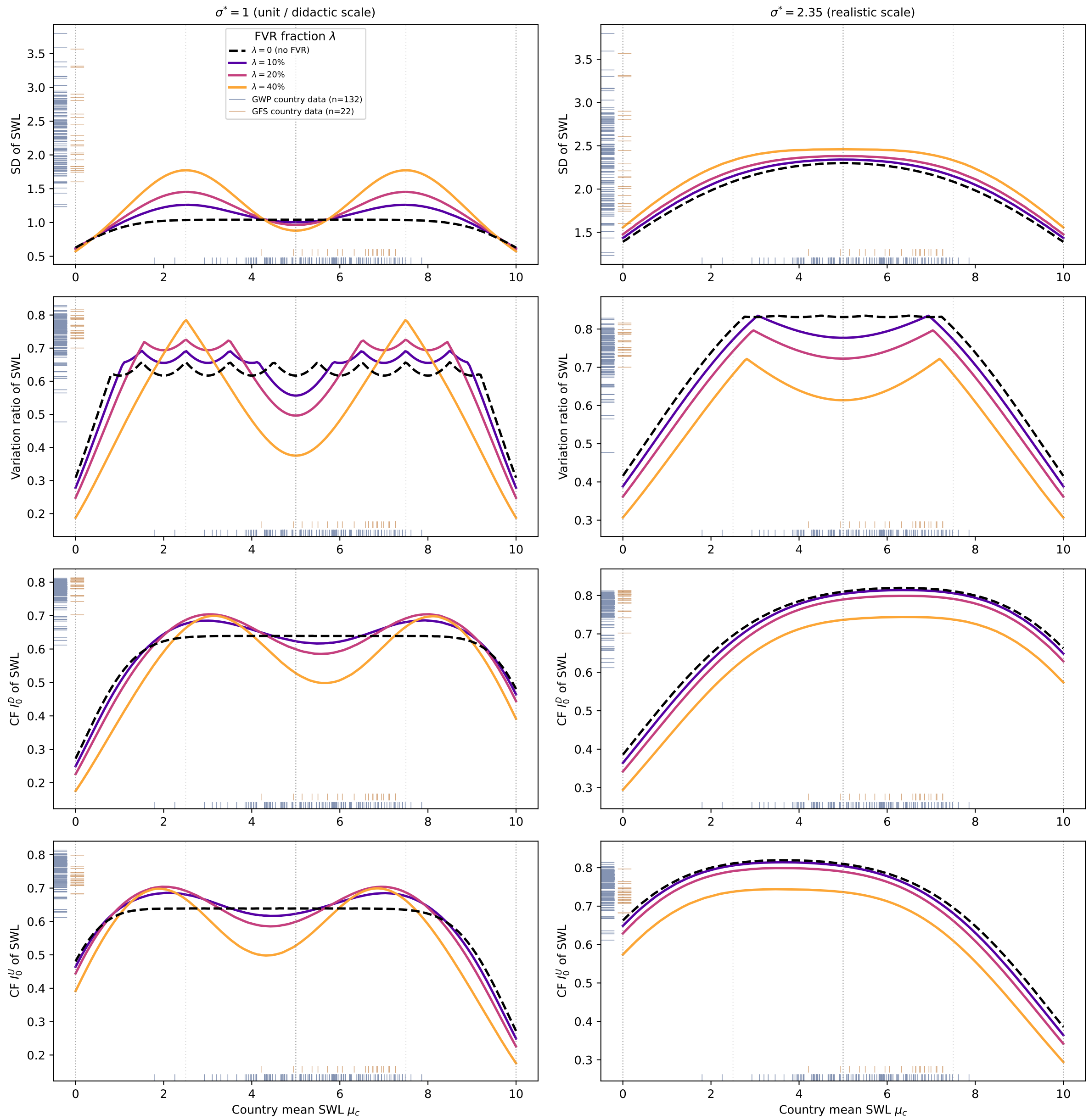

Figure 1: Analytic FVR bias curves for four dispersion statistics. Analytic FVR bias in four dispersion statistics as a function of the sample mean μ, one curve per focal-responding share $\lambda \in \{0, 0.1, 0.2, 0.4\}$; there is no sampling noise (Appendix C.1). **Rows:** standard deviation; the ordinal variation ratio ($1 - \max_k p_k$, the share of responses not at the modal category); the Cowell–Flachaire downward-looking status index I_0^D; and the Cowell–Flachaire upward-looking index I_0^U. **Columns:** the unit scale $\sigma^* = 1$ (left) and the realistic scale $\sigma^* = 2.35$ (right) — the latter the empirical mean within-country life satisfaction standard deviation across the GWP and GFS countries. In each panel the λ=0 curve is the bounded-scale (censoring) baseline; its vertical distance to the λ>0 curves is the FVR bias. Marginal rugs mark the real GWP+GFS country means (horizontal axis) and country dispersion values (vertical axis). The λ legend (top-left panel) is common to all eight panels.

mode switches to the next category — a discreteness artefact that appears even on an unbounded ordinal scale and has nothing to do with censoring or FVR.

4.3 At realistic dispersion, censoring dominates and rounding recedes

The unit scale of the left column somewhat isolates the rounding mechanism, but it is likely not a good description of empirical data. The within-country standard deviation of raw (reported) life satisfaction across the surveys treated below has average $\sigma^* \approx 2.35$, and at that dispersion (Figure 1, right column) the emphasis inverts.

First, the FVR bias *washes out*. Its peak in the standard deviation falls from 0.74 at unit dispersion to 0.25; the variation ratio's peak falls from 0.25 to 0.22; and the Cowell–Flachaire indices fall from 0.14 (I_0^D) and 0.14 (I_0^U) to 0.10 and 0.10 respectively. A wider latent distribution straddles several focal neighbourhoods whatever the mean, smearing away the mean-on-focal structure that drives the swing at $\sigma^* = 1$.

Second, and conversely, the *bounded-scale* bias grows and now reaches every statistic, ordinal or not. The standard deviation's λ=0 (censoring-only) range across the full scale widens from 0.41 to 0.91, and the variation ratio's from 0.35 to 0.42 (likewise the Cowell–Flachaire indices: I_0^D from 0.37 to 0.43 and I_0^U from 0.37 to 0.43; Figure 1, right column). More telling than the aggregate range: at this wider dispersion the tails reach the 0/10 boundary from means well inside where empirical country-level data sit — visibly, in the right-column variation-ratio panel, the λ=0 curve is now bowed well short of the endpoints, not only at them. The clean ordinal isolation of Section 4.2 is thus a feature of the unit latent scale: at the empirically-motivated level of dispersion, every measure, ordinal or not, acquires a mean-dependent censoring bias.

Taken together, the two columns say that, at the scale country data occupies, the dominant contaminant of the standard deviation is bounded-scale censoring, with FVR a smaller perturbation on top of it.[6] This anticipates the empirical finding of Section 5, where a substantial bounded-scale structure in the mean–dispersion association coexists with a more modest *net* FVR correction. The unit-scale column identifies the rounding mechanism; the realistic column locates its magnitude relative to censoring.

4.4 GHM's robustness defences do not neutralize the contamination

The robustness arguments this literature deploys are only partly protective. Goff et al. (2018) defend their welfare interpretation of the mean–dispersion association with three: an allowance for bounded-scale reporting, a purely ordinal dispersion measure (the variation ratio, the share of responses away from the mode), and the observation that the association is stronger among respondents who profess to care most about inequality. None of the three neutralizes FVR.

Consider the ordinal-measure defence first. As the variation-ratio row of Figure 1 shows (Section 4.2 and 4.3), the reported variation ratio swings with the country mean by up to 0.25 at

[6] It should be noted, however, that FVR's relationship to correlates like education may still induce deeper biases.

	(1)	(2)
Reported SD σ_c^{obs}	**4.3**†	**4.4**†
	(.026)	(.026)
Inequality aversion a	**−.084**†	**.33**†
	(.002)	(.029)
$a \times \sigma_c^{\text{obs}}$ ($\beta_{e\sigma}$)		**−.38**†
		(.026)
log income	−.002	−.002
	(.004)	(.004)
Constant	**1.22**†	**1.14**†
	(.029)	(.029)

Significance: **0.1%**† **1%*** *4%* *10%*$^+$

Table 1: GHM Test 1 falsification regression. OLS of reported life satisfaction h_{ij} on the country reported SD σ_c^{obs}, simulated inequality aversion a, their interaction (column 2), and log income, in a synthetic cross-section with *no* genuine wellbeing–dispersion channel. Column (2) adds the aversion × SD interaction: $\beta_{e\sigma} = -0.38$ ($p < 0.001$) carries GHM's Test-1 sign despite the absence of any real inequality channel — the within-individual artefact visualised in Figure D1. $N = 225{,}000$ individuals; $R^2 = 0.118$ (1), 0.119 (2). Standard errors below estimates.

unit dispersion, as λ rises, and at the realistic $\sigma^* = 2.35$ the measure acquires a genuine censoring dependence as well.

The bounded-scale allowance, GHM's first defence, corrects only for a censoring channel that is common to all respondents; FVR, by contrast, is a reporting behaviour that differs across respondents — concentrated among the less educated (Barrington-Leigh, 2024). Thus, a bounded-scale correction leaves the FVR contamination untouched.

GHM's third defence, that the mean–dispersion association is stronger among respondents who profess to care most about inequality[7] (GHM's Test 1), fares no better in principle. Assuming only that lower-educated respondents exhibit higher FVR rates and that the professed inequality-aversion attitude GHM's Test 1 relies on is itself correlated with education ($\rho_a = -0.5$; see Appendix C.4), the same interaction they report arises spuriously, with no genuine inequality channel in effect. To illustrate this, a regression on synthetic data in which this attitude has no impact is shown in Table 1. The relevant interaction coefficient $\beta_{e\sigma} = -0.38$ ($p < 0.001$) is the same sign as found by GHM, who interpret it as evidence that the inequality-averse perceive real inequality. In this synthetic scenario, the effect operates *within* countries through education being linked at the individual level both to inequality aversion and to FVR. Appendix D unpacks this coefficient decile by decile and shows, via a country-level split, that the artefact is indeed carried within countries rather than between them.

The theory and simulations establish what *could* happen. Whether it does — whether realized FVR contamination disrupts rankings or materially distorts the observed mean–dispersion association — is an empirical question, treated next.

[7] In most countries of the World Values Survey, higher education is correlated with higher tolerance of inequality according to the measure used by GHM — as assumed here. However, this question in the WVS is actually phrased around incentives and redistribution, not about concern for inequality (or poverty). These preferences are generally treated as distinct (Bénabou and Tirole, 2006; Alesina and La Ferrara, 2005).

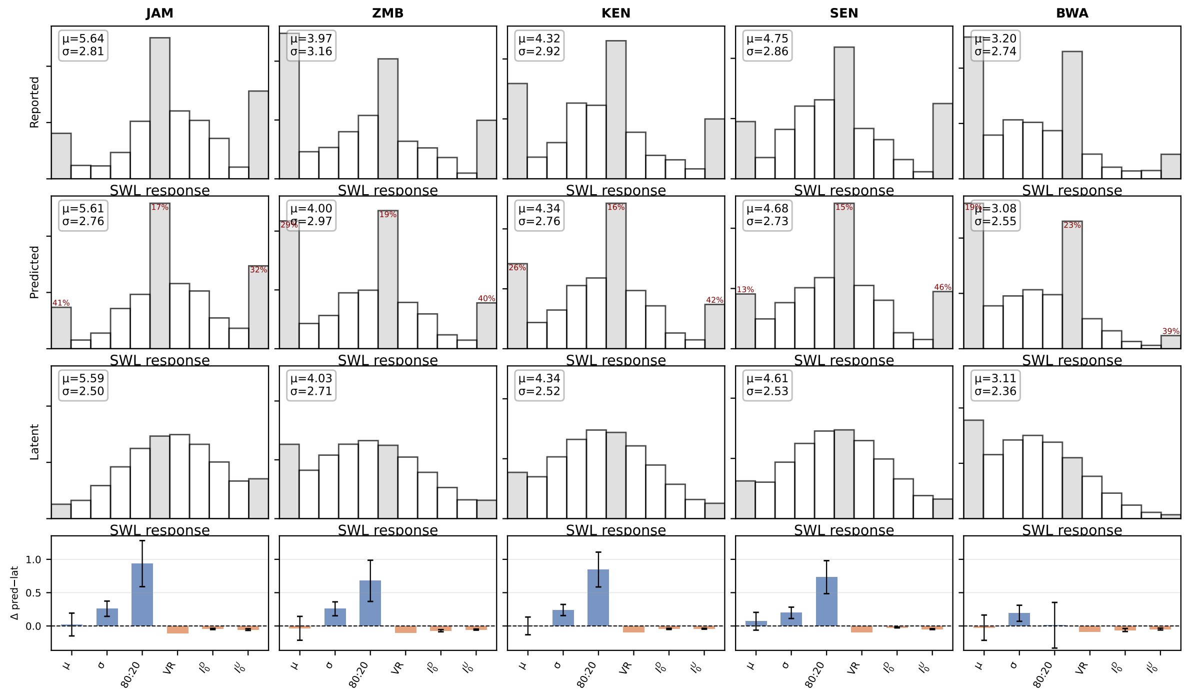

Figure 2: Predicted and latent PMFs, largest-correction countries. Predicted (with FVR; top) and latent (corrected; middle) per-country PMFs for the countries with the largest $|\Delta\sigma|$, Gallup Cantril ladder, with per-country differences (predicted − latent) in the mean and in all five featured statistics — σ, the 80:20 difference, the variation ratio, and the Cowell–Flachaire indices (bottom strip).

5 Global estimates

5.1 Data and estimation

The primary dataset is the Gallup World Poll Cantril ladder ("life today"), a repeated cross-section pooled over 2020–2022: $N = 300{,}137$ respondents in 135 countries. Three other datasets are treated in the appendices and replicate the primary results. These are the Gallup World Poll satisfaction-with-life item (2007–2010; $N = 109{,}375$, 114 countries; Appendix H), which comes from a different period, and the Global Flourishing Study life-today and satisfaction-with-life items, asked of the same respondents in a single cross-section ($N = 177{,}353$ and $177{,}157$; 22 countries; Appendix I).

The reallocation model is fitted separately to each dataset. For each country I extract the posterior predicted and latent PMFs and compute the statistics of Section 2 on each. Credible intervals for every cross-country quantity are computed by re-evaluating each within each of 200 posterior draws.

5.2 What the distributions look like

Figure 2 shows reported, predicted, and latent PMFs for the countries with the largest corrections to the standard deviation. The distributions differ across countries in modes, skewness, and extreme-category mass in ways that no scalar represents. Figure 3 shows the opposite extreme — the countries

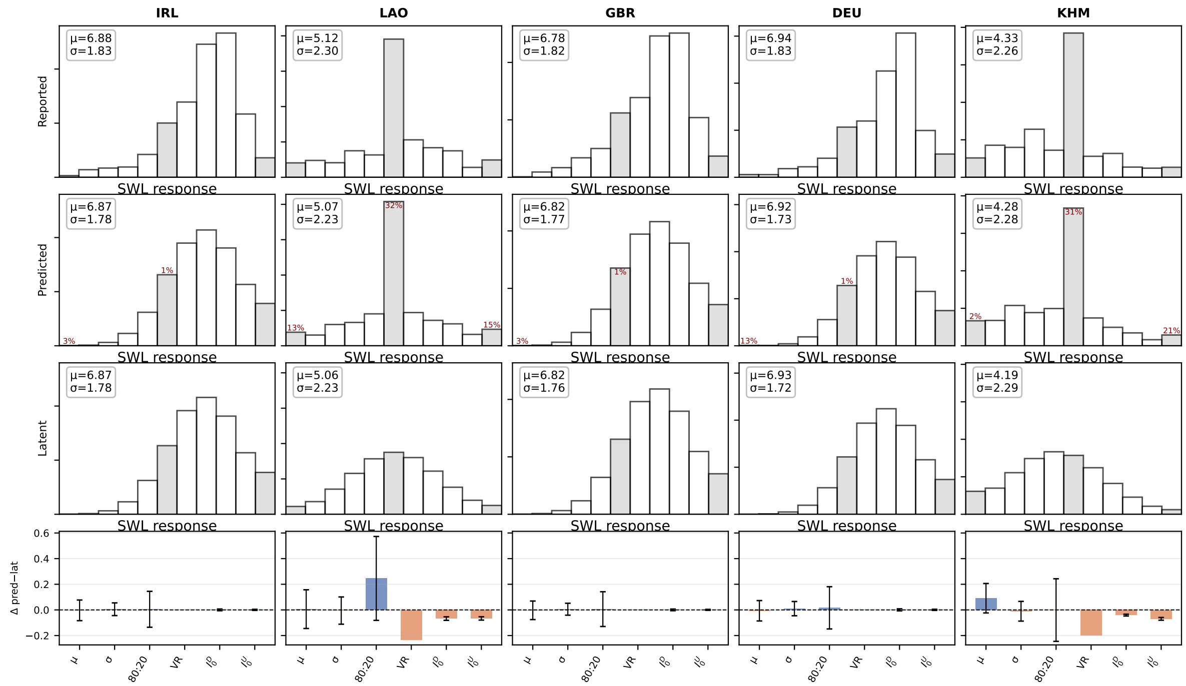

Figure 3: Predicted and latent PMFs, smallest-correction countries. As Figure 2, but for the countries with the *smallest* $|\Delta\sigma|$ — those where FVR barely moves the σ, though that is not necessarily true of the scalar measures shown in the bottom strip.

where FVR barely moves the dispersion — for which the predicted and latent PMFs nearly coincide. The model allows for two distinct sources of endpoint mass at 0 and 10. One is boundedness: a latent distribution wide enough to spill past the scale's endpoints piles up at 0 or 10 under the ordered-probit censoring, the same mechanism as in an OLS model. The other is FVR: probability mass from the neighbourhood of an endpoint is reallocated to the endpoint itself. Countries with broad latent distributions therefore retain substantial endpoint mass in $\hat{\boldsymbol{\pi}}_c$. Such cases are rare — Appendix F collects the five countries where the latent distribution retains the most endpoint mass — and their rarity is why the dispersion of the latent PMF tracks the continuous latent scale σ_c so closely (Section 5.5).

5.3 The scale and the sign of the correction

Table 2 reports, for each statistic: the cross-country ranges of the predicted and latent versions, the median absolute correction, the correction as a share of the cross-country range, and the Spearman correlation between predicted and latent country rankings.[8] Three things stand out in the table:

The correction is material, and almost entirely one-signed. FVR inflates the measured standard deviation in 133 of 135 countries (median signed correction $+0.090$; range $[-0.01, +0.26]$,

[8]The table compares the model's predicted with its latent rankings because that difference is attributable to the FVR channel alone. The practitioner-relevant variant — how stable rankings computed from the *raw* reported standard deviation are against the latent ranking — is Spearman $\rho = 0.966$.

Table 2: Scale of the FVR correction, GWP Cantril ladder. Scale of the FVR correction in the Gallup World Poll Cantril ladder ($N = 300{,}137$; 135 countries; specification (A)). Median $|\Delta|$ is the median across countries of $|T(\hat{\mathbf{P}}_c^{\text{pred}}) - T(\hat{\boldsymbol{\pi}}_c)|$; Rank ρ is the Spearman correlation between predicted and latent country rankings.

	Pred range	Latent range	Median $|\Delta|$	Δ / range	Rank ρ
Mean	[1.785, 7.899]	[1.794, 7.926]	0.029	0%	0.999
SD	[1.214, 3.682]	[1.166, 3.617]	0.090	4%	0.992
80:20	[3.424, 9.713]	[3.271, 9.399]	0.222	4%	0.988
CF I_0^D	[0.622, 0.810]	[0.633, 0.834]	0.019	10%	0.738
CF I_0^U	[0.631, 0.816]	[0.648, 0.835]	0.020	11%	0.820
Variation ratio	[0.496, 0.869]	[0.676, 0.880]	0.075	20%	0.016

the largest in Jamaica). That is: in these data, the analytic possibility of sign-switching is barely realized, because the effect of rounding toward the extreme focal values — which stretches the reported distribution — dominates that of rounding toward 5 in nearly every country. Figure 4 shows the corrections against the country mean, coloured by where each country's focal mass sits.

Rankings on the cardinal statistics largely survive. Because the corrections are one-signed and partly common, the country ranking on the standard deviation is nearly preserved: Spearman $\rho = 0.992$ between predicted and latent (96% credible interval [0.989, 0.993]),[9] with the 80:20 difference behaving similarly ($\rho = 0.988$). The mean is essentially exact ($\rho = 0.999$; median $|\Delta| = 0.029$): rounding to 10 raises it and rounding to 0 lowers it, and the two appear coincidentally to cancel each other to a remarkable degree. Offsetting biases, in other words, do much of the work that an optimist would have hoped for — in levels they do not cancel, but in ranks they largely do.

The ordinal indices are the most disturbed. The Cowell–Flachaire downward-looking index ranks at $\rho = 0.738$ (CI [0.688, 0.764]) and the upward-looking at $\rho = 0.820$ (CI [0.778, 0.835]), on the indices that the ordinal literature advocates precisely for their robustness properties. The mechanism is the one identified in Section 4: these indices are built from the cumulative response shares at each category, and FVR deforms those shares precisely at the most-used, focal categories.

The variation ratio fares worse still: $\rho = 0.016$ (CI [-0.035, 0.126]) — statistically indistinguishable from zero, the lowest rank stability of any of these statistics. This is the real-data counterpart of the synthetic result of Section 4.4: being ordinal buys the variation ratio no protection, because FVR routinely relocates a country's modal response category rather than merely reshaping the mass around it.

The countries where the correction bites hardest, and those where it barely registers, are listed in Table 3. The magnitude is substantial when read as a focal-value response index (FVRI): across

[9] The 96% level is my chosen convention throughout: credible intervals are per-draw posterior percentile intervals $[2, 98]$. Note also that Spearman point estimates are correlations of the posterior-mean measures, while the credible intervals are computed from per-draw correlations, whose mean differs slightly from the point estimate (a Jensen gap); an interval therefore need not be centred on the quoted ρ.

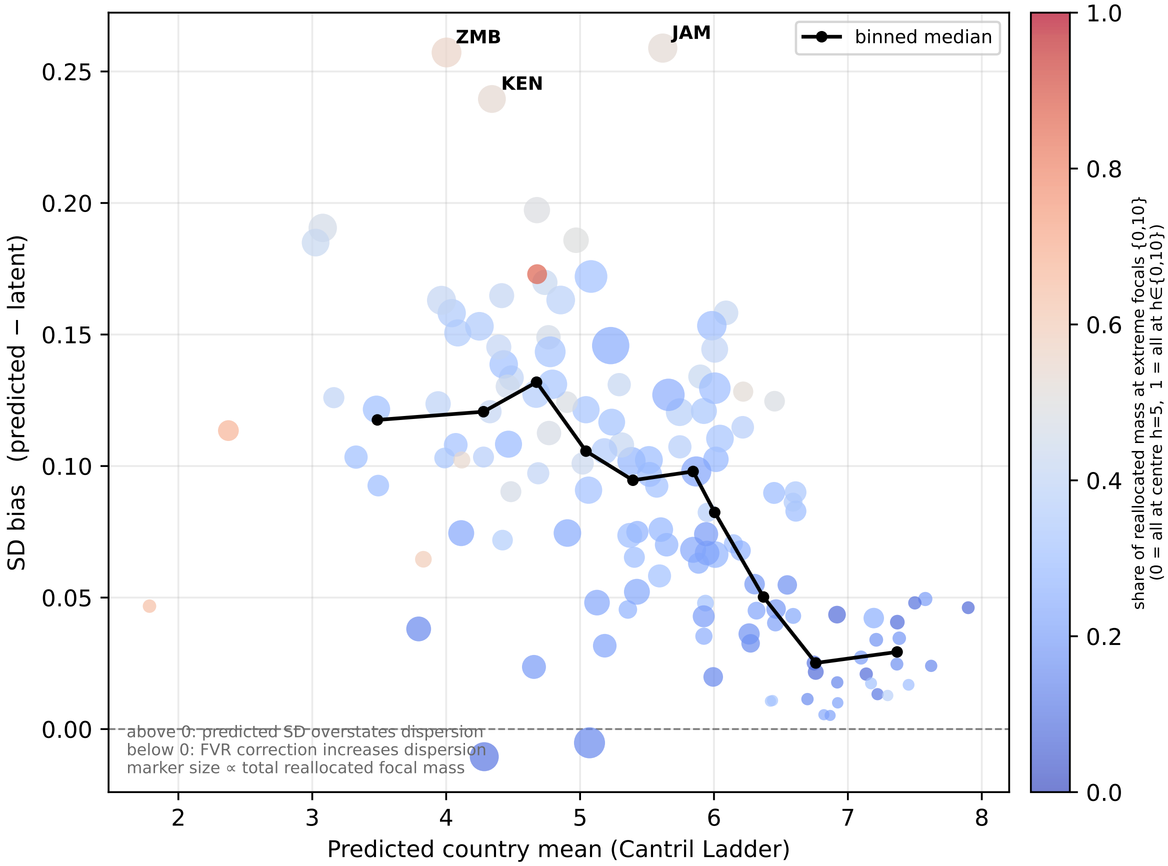

Figure 4: Standard-deviation correction versus country mean. Standard-deviation correction (predicted − latent) versus predicted country mean, Gallup Cantril ladder. Colour: share of FVR mass at the extreme focal values $\{0, 10\}$ relative to the central focal 5; marker size: total reallocated mass.

Table 3: Largest and smallest standard-deviation corrections by country. Countries with the largest and smallest standard-deviation corrections (signed $\Delta\sigma$ = predicted − latent), top and bottom of the 135 Gallup Cantril-ladder countries, with the per-focal-value FVRI (the excess focal mass relocated by FVR, predicted minus latent, as a share of each neighbourhood's predicted mass); note that because each FVRI is a share of a possibly tiny neighbourhood mass, a large percentage can coexist with negligible absolute reallocation — e.g., FVRI_0 in high-mean countries with little low-end mass. Specification (A).

Country	Pred SD	Latent SD	ΔSD	FVRI_0	FVRI_5	FVRI_{10}
Jamaica	2.757	2.498	+0.259	41%	17%	32%
Zambia	2.962	2.704	+0.257	29%	19%	40%
Kenya	2.762	2.522	+0.240	26%	16%	42%
Senegal	2.728	2.530	+0.197	13%	15%	46%
Botswana	2.543	2.352	+0.191	19%	23%	39%
Cote d'Ivoire	2.631	2.445	+0.186	19%	13%	35%
Zimbabwe	2.389	2.204	+0.185	17%	22%	45%
Mauritania	3.099	2.926	+0.173	8%	2%	32%
France	1.678	1.667	+0.011	8%	2%	1%
Spain	1.713	1.702	+0.011	12%	1%	0%
Italy	1.649	1.639	+0.011	14%	1%	0%
Germany	1.730	1.720	+0.010	13%	1%	0%
United Kingdom	1.765	1.760	+0.005	3%	1%	0%
Ireland	1.783	1.778	+0.005	3%	1%	0%
Lao PDR	2.223	2.228	-0.005	13%	32%	15%
Cambodia	2.281	2.291	-0.010	2%	31%	21%

the 135 countries the median country relocates onto its focal values, through FVR, a fraction 0.174 of its low-end neighbourhood mass (onto 0), 0.133 of its central-neighbourhood mass (onto 5), and 0.139 (onto 10) — where each fraction is the excess focal mass (predicted minus latent) as a share of that neighbourhood's predicted mass — for an overall relocated focal mass of 0.137 in the median country. Across countries this overall relocated mass is largest in Tajikistan.

A caveat on magnitudes: the comparison above is between the model's predicted and latent distributions, which isolates the FVR channel. The gap between the *raw* reported statistics and the latent ones is wider, but the additional spread reflects model lack-of-fit rather than FVR; conflating the two would overstate the correction. How much of the relocated focal mass translates into a net correction — and how much cancels within a country — is shown statistic by statistic in Appendix G: the net correction tracks the total FVRI closely for the dispersion and ordinal statistics, but hardly at all for the mean, where rounding toward 10 and toward 0 largely offset.

5.4 The GHM regression, before and after correction

Goff et al. (2018) regress country-mean life satisfaction on its standard deviation, conditional on log GDP per capita and (in some specifications) the income Gini. Table 4 re-runs the country-level regression on the Gallup Cantril ladder with three versions of the dispersion regressor — raw reported,

Table 4: GHM-style regression on raw, predicted, and latent σ. GHM-style country-level regression of mean life satisfaction on σ of life satisfaction, log GDP per capita, and income Gini, using raw, predicted, and latent σ as alternative regressors. Gallup Cantril ladder, 133 countries; WLS by country sample size. "Implied ΔM over σ IQR" rescales each coefficient by the interquartile range of its own regressor. Standard errors, in parentheses, are from re-estimating the regression within each posterior draw.

	raw			pred			latent		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$\hat{\sigma}$ raw	***−.43***		**−.90**†						
	(0.189)		(0.203)						
Implied ΔM over σ IQR	***−.33***		**−.65**†						
	(0.145)		(0.145)						
$\hat{\sigma}$ predicted				***−.44***		**−.86**†			
				(0.194)		(0.206)			
Implied ΔM over σ IQR				***−.30***		**−.54**†			
				(0.132)		(0.131)			
$\hat{\sigma}$ latent							*−.41*$^+$		**−.82**†
							(0.206)		(0.218)
Implied ΔM over σ IQR							*−.25*$^+$		**−.47**†
							(0.127)		(0.124)
Income Gini		.008	*.013*$^+$		.008	.012		.008	.011
		(0.008)	(0.008)		(0.008)	(0.008)		(0.008)	(0.008)
Log GDP PPP	**.72**†	**.88**†	**.57**†	**.73**†	**.88**†	**.61**†	**.75**†	**.88**†	**.65**†
	(0.082)	(0.059)	(0.087)	(0.081)	(0.059)	(0.085)	(0.079)	(0.059)	(0.083)
Constant	−.49	**−3.3**†	1.64	−.59	**−3.3**†	1.16	−.89	**−3.3**†	.66
	(1.171)	(0.703)	(1.282)	(1.146)	(0.703)	(1.251)	(1.135)	(0.703)	(1.239)
N	133	121	121	133	121	121	133	121	121
R^2	0.659	0.664	0.712	0.658	0.664	0.707	0.655	0.664	0.700

Significance: **0.1%**† **1%*** ***4%*** *10%*$^+$

model-predicted, and latent — weighted by country sample size.

With the income-Gini control, the coefficient on the raw σ is -0.90 $[-1.33, -0.48]$, on the predicted σ -0.86 $[-0.92, -0.72]$, and on the latent σ -0.82 $[-0.90, -0.68]$, where each bracketed range is a 96% credible interval.[10] Without the Gini control the corresponding values are about half as much, with the latent interval $[-0.48, -0.30]$. The cross-sectional GHM association thus survives the FVR correction in these data: the latent-σ coefficient is modestly attenuated relative to the raw one, retains its sign, and remains conventionally significant. This is an empirical outcome rather than a foregone conclusion — Section 4 shows that FVR contaminates cross-country dispersion comparisons in a mean-dependent way that could have attenuated, inflated, or reversed the association — and which of these occurs could only be established by estimating the correction.

Two qualifications keep the result in context. First, the deeper fragility of the GHM association documented by Grimes et al. (2023) is logically prior to mine and is untouched by it: in their

[10] For the predicted and latent σ the interval is the posterior credible interval from re-estimating the regression within each posterior draw; the raw σ is a fixed function of the data, so its interval is the classical WLS confidence interval.

replication the SD coefficient loses significance once country fixed effects are included and is "virtually zero" with wave dummies added, and their theoretically preferred ordinal indices yield significant associations of *opposite* signs for upward- versus downward-looking status. A cross-sectional coefficient that survives a measurement correction but not a fixed-effects specification, or whose sign depends on the choice among defensible measures, is not yet a foundation for welfare conclusions. Second, my correction speaks only to the FVR channel; the cardinality and boundedness critiques of using the SD at all (Schröderg and Yitzhaki, 2017; Delhey and Kohler, 2011) stand independently.

This sensitivity to the income-Gini control has no counterpart in Goff et al. (2018): in their individual-level regressions the standardized coefficient on the life satisfaction SD is essentially identical with and without the Gini — -0.10 in both columns for the Gallup World Poll — and it is instead the Gini coefficient that shrinks toward insignificance when both enter, a pattern they read as life satisfaction dispersion subsuming the information in income inequality. In my country-level cross-section the dependence runs the other way.

5.5 The latent scale σ_c: dispersion free of both rounding and censoring

Every statistic considered so far, including the latent versions, is a functional of a PMF on the 0–10 categories, and therefore still carries the bounded-scale censoring that Section 4.3 identifies as the dominant contaminant at realistic dispersion. The model, however, delivers one dispersion object that is free of *both* distortions: the latent scale σ_c of equation (1) — the dispersion that Section 3 declares this paper's target, measured before the scale's boundedness or any rounding intervenes. Across the 135 Gallup Cantril-ladder countries, σ_c ranges from 1.07 to 5.58, with median 2.15. Its rank agreement with the latent-PMF standard deviation is Spearman $\rho = 0.985$; the distance of that correlation from one measures how much undoing the censoring, on top of FVR, reorders countries.

Re-running the GHM regression of Section 5.4 with σ_c as the dispersion regressor gives, with the income-Gini control, a coefficient of -0.43 $[-0.49, -0.36]$, and without it -0.17 $[-0.23, -0.13]$, the bracketed ranges again being 96% posterior credible intervals from per-draw re-estimation. This is the mean–dispersion association evaluated in the fully corrected metric. Set against the raw, predicted, and latent-PMF rows of Table 4, these coefficients show directly how much of the association survives once both reporting layers — rounding and censoring — are removed from the dispersion regressor, and they suggest a much weaker relationship. Goff et al. (2018) confronted the same censoring concern by fitting a symmetric logistic latent distribution in each country–wave cluster and re-estimating in "actual" life satisfaction space; their SD coefficient shrank by roughly a fifth to two fifths depending on the survey — from -0.10 to -0.08 in the Gallup World Poll — which they summarize as at most a third of the reported-space association being a reporting artifact. Their adjustment addresses censoring under an assumed symmetric latent shape but leaves the rounding channel untouched, so the σ_c comparison here, which removes both layers, is the more demanding test.

5.6 Dominance comparisons on the corrected distributions

If no scalar is to be trusted, what comparisons remain? Jenkins (2021) shows that non-intersection of generalized Lorenz curves of status is equivalent to a unanimous ranking by the whole Cowell–Flachaire index family, and finds rankable pairs in his six-country application to reported life satisfaction data. I apply the same criteria to the corrected distributions: for every pair of countries I test first-order dominance of the response CDFs and generalized-Lorenz-of-status dominance, on both the predicted and the latent PMFs. Of 9,045 country pairs, 4,773 (53%) are first-order ranked on the predicted distributions and 6,044 (67%) on the latent ones, with 0 reversals among pairs ranked in both; the corresponding GL-of-status counts are 516 and 942; the number of reversals among pairs GL-ranked in both is 1. That GL-of-status dominance certifies far fewer pairs than first-order dominance is what the criterion's strength implies: it demands unanimity over the entire Cowell–Flachaire index family, and because mean status levels differ across countries, the generalized Lorenz curves of status cross easily even where the response CDFs do not. The correction, by removing spikes that cause CDFs to cross, makes the distributions *more* often comparable — and where dominance holds on the latent distributions, the comparison is robust to every choice of index that the dominance theorem covers.

6 Discussion: what survives measurement?

The two claims under test. On the first claim — that cross-country variation in a dispersion statistic of reported life satisfaction reflects underlying wellbeing inequality — the answer assembled here is unfavourable, on three converging grounds: the mean-linked or threshold-saturated statistics (Gini, CV, the below-5 share) confuse two concepts usually intended to be separated (Section 2.2); the ordinal alternatives — the indices designed to repair cardinality, and GHM's own variation-ratio defence — are the statistics most disturbed by FVR (Section 5); and the standard deviation, while rank-stable under the FVR correction, measures a reported distribution whose levels the correction moves in nearly every country, on top of the boundedness and cardinality objections that precede this paper. The $T(\hat{\boldsymbol{\pi}}_c)$ comparisons are censoring-inclusive by design — computed on the printed 0–10 scale so as to remain comparable with reported-scale statistics — while the latent scale σ_c (Section 5.5) is the object corrected for both rounding and censoring. On the second claim — the welfare reading of the mean–dispersion association — the FVR correction is, perhaps surprisingly, exculpatory: the association is not a FVR artifact. It remains, per Grimes et al. (2023), a cross-sectional association that does not survive fixed effects and whose sign is measure-dependent in the ordinal family.

Limitations. A main limitation to this work is that the correction is, inevitably, model-dependent: the cardinal cutpoints are not the most flexible parameterization conceivable for the latent response, but were required for identifiability in the reassignment model — and, as Section 3.3 notes, they presume the equal-spacing cardinality that the relabelling critiques contest, so the ordinal-index conclusions are conditional on that identifying restriction. The model captures boundedness, FVR, country-level unexplained variance, and response shape driven by the distribution of explanatory factors X; it does not accommodate response scale stretching.

7 Conclusion

The contributions of this paper are to (i) identify focal-value rounding as a measurement problem for life satisfaction inequality that is distinct from the ordinal-data critique; (ii) bound its possible effect on the standard deviation at $2\sqrt{\lambda}$ and show in simulation that, with the genuine inequality channel switched off by assumption, FVR contaminates cross-country dispersion comparisons in a mean-dependent way and spuriously reproduces GHM's inequality-aversion-interaction test; (iii) estimate the realized contamination at global scale, applying a FVR correction to cross-country life satisfaction distributions, calculated with full posterior uncertainty; and (iv) show what the correction does to the statistics in use: levels move in nearly every country, cardinal rankings largely survive, the ordinal indices are the most disturbed, and the Goff et al. association survives with modest attenuation while remaining fragile on grounds others have established.

My constructive conclusion is not to suggest or seek a better scalar. It is that the distribution — or a corrected version of it — is the right object of analysis. Dominance comparisons on latent distributions resolve more country pairs than on reported ones and are robust across the entire index family they cover; distributional plots carry the skewness and extreme-mass features that scalars discard; and targeted group comparisons answer policy questions that no population-wide index addresses.

References

Abul Naga, Ramses H. and Tammy Yalcin (2008). “Inequality measurement for ordered response health data”. In: *Journal of Health Economics* 27.6, pp. 1614–1625. DOI: 10.1016/j.jhealeco.2008.07.015.

Alesina, Alberto and Eliana La Ferrara (2005). “Preferences for Redistribution in the Land of Opportunities”. In: *Journal of Public Economics* 89.5–6, pp. 897–931. DOI: 10.1016/j.jpubeco.2004.05.009.

Allison, R. Andrew and James E. Foster (2004). “Measuring health inequality using qualitative data”. In: *Journal of Health Economics* 23.3, pp. 505–524. DOI: 10.1016/j.jhealeco.2003.10.006.

Apouey, Bénédicte (2007). “Measuring health polarization with self-assessed health data”. In: *Health Economics* 16.9, pp. 875–894. DOI: 10.1002/hec.1284.

Barrington-Leigh, C.P. (Feb. 2024). “The econometrics of happiness: Are we underestimating the returns to education and income?” In: *Journal of Public Economics* 230. ISSN: 0047-2727. DOI: https://doi.org/10.1016/j.jpubeco.2023.105052.

— (Jan. 2026). *Are international happiness rankings reliable?* Tech. rep. URL: https://alum.mit.edu/www/cpbl/publications/Barrington-Leigh-DRAFT2025-global-differences-CLE.pdf.

Bénabou, Roland and Jean Tirole (2006). “Belief in a Just World and Redistributive Politics”. In: *The Quarterly Journal of Economics* 121.2, pp. 699–746. DOI: 10.1162/qjec.2006.121.2.699.

Cowell, Frank A. and Emmanuel Flachaire (2017). “Inequality with Ordinal Data”. In: *Economica* 84.334, pp. 290–321. DOI: 10.1111/ecca.12232.

Delhey, Jan and Ulrich Kohler (2011). “Is happiness inequality immune to income inequality? New evidence through instrument-effect-corrected standard deviations”. In: *Social Science Research* 40.3, pp. 742–756. DOI: 10.1016/j.ssresearch.2010.11.004.

Ferrer-i-Carbonell, A. and P. Frijters (2004). “How Important is Methodology for the estimates of the determinants of Happiness?” In: *The Economic Journal* 114.497, pp. 641–659. DOI: 10.1111/j.1468-0297.2004.00235.x.

Goff, Leonard, John F Helliwell, and Guy Mayraz (2018). “Inequality of Subjective Well-being as a Comprehensive Measure of Inequality”. In: *Economic Inquiry* 56.4, pp. 2177–2194. DOI: 10.1111/ecin.12582.

Gravel, Nicolas, Brice Magdalou, and Patrick Moyes (2021). “Ranking distributions of an ordinal variable”. In: *Economic Theory* 71.1, pp. 33–80. DOI: 10.1007/s00199-020-01262-4.

Grimes, Arthur, Stephen P Jenkins, and Florencia Tranquilli (2023). “The relationship between subjective wellbeing and subjective wellbeing inequality: An important role for skewness”. In: *Journal of Happiness Studies* 24.1, pp. 309–330. DOI: 10.1007/s10902-022-00591-6.

Hasegawa, Hikaru and Kazuhiro Ueda (2011). “Measuring inequality of subjective well-being: A Bayesian approach”. In: *The Journal of Socio-Economics* 40.5, pp. 700–708. ISSN: 1053-5357. DOI: 10.1016/j.socec.2011.05.009. URL: https://www.sciencedirect.com/science/article/pii/S1053535711000709.

Jenkins, Stephen P. (2021). “Inequality Comparisons with Ordinal Data”. In: *Review of Income and Wealth* 67.3, pp. 547–563. DOI: 10.1111/roiw.12489.

Kaiser, Caspar F and Maarten Vendrik (2020). *How threatening are transformations of happiness scales to subjective wellbeing research?* Tech. rep. IZA Discussion Papers.

Kalmijn, Wim and Ruut Veenhoven (2005). “Measuring Inequality of Happiness in Nations: In Search for Proper Statistics”. In: *Journal of Happiness Studies* 6 (4), pp. 357–396. ISSN: 1389-4978. DOI: 10.1007/s10902-005-8855-7.

OECD (2025). *OECD Guidelines on Measuring Subjective Well-being (2025 Update)*. Tech. rep. Paris. DOI: 10.1787/9203632a-en. URL: https://www.oecd.org/en/publications/oecd-guidelines-on-measuring-subjective-well-being-2025-update_9203632a-en.html.

Panasiuk, Sofia L., Anthony McCanny, and Felix Cheung (2025). “Methods Reflect Values: Evaluating the Shortcomings of the Average for Measuring Population Well-Being”. In: *Journal of Personality and Social Psychology* 128.6, pp. 1355–1370. DOI: 10.1037/pspp0000549.

Permanyer, Inaki, Suman Seth, and Gaston Yalonetzky (2025). “Inequality measurement for bounded variables”. In: *Health Economics* 34.8, pp. 1443–1460.

Schröderg, Carsten and Shlomo Yitzhaki (2017). “Revisiting the evidence for cardinal treatment of ordinal variables”. In: *European Economic Review* 92, pp. 337–358. ISSN: 0014-2921. DOI: https://doi.org/10.1016/j.euroecorev.2016.12.011. URL: http://www.sciencedirect.com/science/article/pii/S0014292117300016.

Sheppard, William Fleetwood (1897). “On the Calculation of the most Probable Values of Frequency-Constants, for Data arranged according to Equidistant Division of a Scale”. In: *Proceedings of the London Mathematical Society* s1-29.1, pp. 353–380. DOI: 10.1112/plms/s1-29.1.353.

Stevenson, Betsey and Justin Wolfers (Aug. 2008). “Happiness Inequality in the United States”. In: Working Paper Series 14220. URL: http://www.nber.org/papers/w14220.

What Works Centre for Wellbeing (2018). *Measuring Wellbeing Inequalities: A How-To Guide*. London: What Works Centre for Wellbeing. URL: https://whatworkswellbeing.org/.

Supplementary material

A Notation

Symbol	Meaning
y_i	Reported life satisfaction of respondent i, integer $\in \{0, \ldots, 10\}$.
$\boldsymbol{\pi}_c$	Latent (full-scale) PMF for country c.
$\mathbf{P}_c^{\text{obs}}$	Raw empirical PMF for country c.
$\hat{\mathbf{P}}_c^{\text{pred}}$	Posterior-mean predicted reported PMF.
$\hat{\boldsymbol{\pi}}_c$	Posterior-mean latent PMF for country c.
F_c	Response CDF for country c.
s_i	Cowell–Flachaire status, $s_i = F(u_i)$ (downward) or $1 - F(u_i^-)$ (upward).
$f \in \{0, 5, 10\}$	Focal value; λ the focal-responding share.
X_S	Latent-layer covariates (constant, ln HH income, education, gender).
X_N	Rounding-layer covariates (constant, education).
$c_k = k - 5.5$	Cardinal cutpoints, $k = 1, \ldots, 10$ ($c_0 = -\infty$, $c_{11} = +\infty$).
σ_c	Latent scale of country c (hierarchical).
$\sigma, \sigma^{\text{obs}}, \hat{\sigma}^{\text{pred}}, \sigma^*$	SD: generic, reported, predicted, full-scale (latent).
I_α^D, I_α^U	Cowell–Flachaire downward-/upward-looking ordinal indices.

B Proof of Proposition 1

Let T_i^{full} be the full-scale integer report of individual i and T_i the actual report; for non-FVR respondents these coincide, while for FVR respondents $T_i \in \{0, 5, 10\}$. Define $D_i = T_i - T_i^{\text{full}}$, so $D_i = 0$ for non-FVR respondents.

For an FVR respondent, T_i^{full} is the nearest integer to her latent assessment and T_i the nearest focal value. The maximum discrepancy is 2, occurring near the rounding boundaries: a latent value just below 2.5 gives $T^{\text{full}} = 2$, $T = 0$ ($D = -2$); just above 2.5 gives $T^{\text{full}} = 3$, $T = 5$ ($D = +2$); and analogously near 7.5. Hence $|D_i| \leq 2$ for all respondents, and

$$\sigma_D^2 = \text{Var}(D) \ \leq \ E[D^2] \ \leq \ 4\lambda,$$

so $\sigma_D \leq 2\sqrt{\lambda}$. For any random variables X and Y, the L^2 triangle inequality gives $|\sigma_X - \sigma_Y| \leq \sigma_{X-Y}$; applying this with $X = T$ and $Y = T^{\text{full}}$ yields $|\sigma^{\text{obs}} - \sigma^*| \leq \sigma_D \leq 2\sqrt{\lambda}$. The first inequality binds only if $\text{corr}(T, T^{\text{full}}) = 1$, i.e. only if every respondent's (T^{full}, T) pair lies on one common line; this is automatic when $\lambda = 1$, since any two points are collinear, but for $\lambda < 1$ it constrains where the unrounded remainder can sit.

*Tightness, positive direction (*FVR *inflates the SD), for any λ.* Let a $\lambda/2$ share of respondents have latent assessments just below 2.5 ($T^{\text{full}} = 2$, $T = 0$, $D = -2$) and a $\lambda/2$ share just above 7.5

($T^{\text{full}} = 8$, $T = 10$, $D = +2$), both FVR; let the remaining $1 - \lambda$ share sit exactly at the scale midpoint ($T^{\text{full}} = 5$), non-FVR but reporting 5 regardless since $5 \in \{0, 5, 10\}$. Pairing the two extremal points from *opposite* thresholds, rather than two adjacent to the same threshold, puts the midpoint exactly on the line joining them, since 5 is both the arithmetic mean of 2 and 8 and of their images 0 and 10; pairing two points adjacent to the same threshold instead forces the line through them to cross the diagonal at the non-integer $T^{\text{full}} = 2.5$, so no unrounded respondent can join it and that construction attains the bound only at $\lambda = 1$. Then $\sigma^* = 3\sqrt{\lambda}$, $\sigma^{\text{obs}} = 5\sqrt{\lambda}$, and $\sigma^{\text{obs}} - \sigma^* = 2\sqrt{\lambda}$ for every $\lambda \in [0, 1]$.

*Tightness, negative direction (*FVR *compresses the SD), for any* λ. Let a $\lambda/2$ share sit just above 2.5 ($T^{\text{full}} = 3$, $T = 5$, $D = +2$) and a $\lambda/2$ share just below 7.5 ($T^{\text{full}} = 7$, $T = 5$, $D = -2$), both FVR; let the remaining $1 - \lambda$ share again sit at the midpoint ($T^{\text{full}} = 5 = T$). The two FVR branches already share a common rounded image, 5, so the midpoint mass leaves $T \equiv 5$ identically. Then $\sigma^* = 2\sqrt{\lambda}$, $\sigma^{\text{obs}} = 0$, and $\sigma^* - \sigma^{\text{obs}} = 2\sqrt{\lambda}$ for every $\lambda \in [0, 1]$. □

C The synthetic data-generating process

This appendix specifies two objects of different kinds. The analytic bias curves of Figures 1 and E1 need only a latent mean, an FVR fraction, and a latent dispersion; their construction is given in Appendix C.1. Everything else here — the individual layer, the country cross-section, and the inequality-aversion attitude — serves *only* the single simulated cross-section behind the GHM Test 1 falsification: the regression of Table 1 and its decile-by-decile unpacking in Figure D1.

C.1 The analytic bias curves

The bias curves of Figure 1 and E1 carry no sampling noise and no country, education, or income structure. For a given latent mean μ and dispersion σ^*, the reported probability mass function is formed directly: a latent $\mathcal{N}(\mu, \sigma^*)$ is clipped and rounded to the 0–10 integers, and then a fraction λ of its mass is moved to the nearest focal value ($\{0, 1, 2\} \to 0$, $\{3, \ldots, 7\} \to 5$, $\{8, 9, 10\} \to 10$). Each dispersion statistic is evaluated on that exact PMF. The curves sweep μ over $[0, 10]$ at $\lambda \in \{0, 0.1, 0.2, 0.4\}$ (up to 40%), at the two latent dispersions $\sigma^* = 1$ and $\sigma^* = 2.35$. The λ=0 curve is the bounded-scale (censoring) baseline; its range over the full sweep measures the censoring bias. The realistic value $\sigma^* = 2.35$ is the empirical pooled mean within-country life satisfaction standard deviation across the GWP and GFS countries.

C.2 The individual layer

A respondent i in country c has a latent wellbeing

$$z_i = m_c + \varepsilon_i, \qquad \varepsilon_i \sim \mathcal{N}(0, 1),$$

where m_c is the country mean centre, so the within-country latent standard deviation is exactly 1 in every country by construction. The full-scale integer report is $y_i^* = \text{clip}\big(\text{round}(z_i),\, 0,\, 10\big)$. Respondent i also carries education $x_i \sim \mathcal{N}(\mu_x, 1)$ and log income $\ln y_i \sim \mathcal{N}(0, \sigma_y^2)$, with the income dispersion σ_y a single value fixed across all countries; neither enters the wellbeing equation above. Education governs the rounding (below); log income has no effect on wellbeing and is retained only as a control in the GHM falsification regression of Section 4, mirroring their specification. Focal-value rounding is then applied with the same neighbourhoods as the estimation model: with probability $p_i = \text{logit}^{-1}(\gamma_0 - \gamma_1 x_i)$ respondent i replaces y_i^* by its nearest focal value ($\{0,1,2\} \to 0$, $\{3,\ldots,7\} \to 5$, $\{8,9,10\} \to 10$); otherwise she reports y_i^*. With $\gamma_1 > 0$ lower-education respondents round more often. At $x = 0$ — the cross-country centre of mean education (Appendix C.3) — the focal share is $\text{logit}^{-1}(\gamma_0)$ (about 0.27 at $\gamma_0 = -1$, 0.12 at $\gamma_0 = -2$); a country with mean education $\overline{x}_c$ has focal share $\text{logit}^{-1}(\gamma_0 - \gamma_1 \overline{x}_c)$ at its mean.

C.3 The country cross-section

For a cross-section of C countries, each country's wellbeing centre and mean education are

$$m_c = \text{clip}\big(\mu_0 + \sigma_{\text{swl}}^{\text{across}}\, u_c,\ 3,\ 9.5\big), \qquad \overline{x}_c = \sigma_x^{\text{across}}\, v_c,$$

where (u_c, v_c) is a standard bivariate normal with correlation $\rho_{xs} = \text{corr}(\overline{x}_c, m_c)$ — the single cross-country education–wellbeing correlation — μ_0 is a fixed global mean (the "base" of Table C1), and $\sigma_{\text{swl}}^{\text{across}}$ and σ_x^{across} are the across-country standard deviations of the two dimensions. The wellbeing centre is thus shifted and clipped while mean education stays centred at zero. Each country's respondents are then drawn from the individual layer above with this country centre m_c and mean education $\mu_x = \overline{x}_c$; education thus varies both within and across countries, while the latent wellbeing dispersion is held at 1 everywhere.

The assumed magnitudes. The latent (focal-free) within-country dispersion is exactly $\text{SD}(\varepsilon) = 1$ by construction, identical across countries up to the mechanical compression of the bounded 0–10 scale near its endpoints. The *assumed* cross-country variation in true wellbeing dispersion is therefore zero: nothing — income, education, or attitude — enters the wellbeing equation, so there is no genuine dispersion channel of any kind. Any cross-country mean–dispersion pattern in the simulation without FVR is thus a bounded-scale censoring artefact, and any further movement once FVR is applied is focal-value rounding — neither is a wellbeing-inequality signal.

C.4 The inequality-aversion attitude

For the GHM Test 1 cross-section each respondent is additionally assigned a simulated inequality-aversion attitude $a_i = \rho_a\, \tilde{x}_i + \sqrt{1-\rho_a^2}\, \nu_i$, $\nu_i \sim \mathcal{N}(0,1)$, with $\tilde{x}_i$ standardised education and $\rho_a = -0.5$ (lower education $\to$ more averse). This correlation is an ingredient of the simulation, not of GHM's argument: it is what lets education-driven FVR impersonate an aversion–dispersion interaction, and its sign is set so the artefact carries the same sign as the interaction GHM report. Crucially a_i enters *no* equation of the DGP: it has no causal effect on wellbeing. The genuine "inequality-averse people perceive real inequality" effect that GHM's Test 1 is designed to detect is therefore assumed to be exactly zero, so any nonzero interaction the test returns is, by construction, a focal-value-rounding artefact and not a signal.

C.5 Parameter settings

Table C1 lists the settings behind the one simulated cross-section, that of the GHM Test 1 falsification (Table 1 and Figure D1). The cross-section is centred at base $= 6.0$ with $C = 150$ countries (matching GHM's cross-country sample size), the regime in which the focal mechanism spans the inflation/deflation boundary.

Table C1: Parameter settings, GHM Test 1 synthetic cross-section. Parameter settings for the GHM Test 1 synthetic cross-section (Figure D1). Nothing enters the wellbeing equation (so true dispersion is fixed), $\gamma_1 = 1$, individual $\sigma_x = 1$, cross-country $\sigma_x^{\text{across}} = 0.7$, fixed log-income dispersion $\sigma_y = 0.6$, and seed 42 (the inequality-aversion attitude uses a secondary seed 9999). "base" is the cross-country mean of m_c; N_c is respondents per country.

Figure	C	N_c	base	$\sigma_{\text{swl}}^{\text{across}}$	γ_0	ρ_{xs}
D1 (inequality aversion)	150	1500	6.0	0.5	-2	0.65

D Decile-level visualization of the synthetic Test 1 artefact

Table 1 reports a single pooled interaction coefficient, $\beta_{e\sigma}$; Figure D1 unpacks that coefficient by estimating the SD–SWL slope separately within each inequality-aversion decile. Two things are visible here that the table alone does not show. First, the decline is monotonic across all ten deciles, not merely consistent with a linear interaction term. Second, and more informative for locating the mechanism, a coarse country-level median split on the same synthetic data yields near-equal subgroup slopes: the artefact is carried *within* countries by individual education rather than by any between-country composition effect, confirming the individual-level channel described in Section 4.4 rather than a spurious cross-country correlation.

E Bias geometry of the excluded statistics

Section 2.2 excludes the Gini coefficient, the coefficient of variation, and the share reporting below 5 from the mean–dispersion analysis, on three related but distinct a priori grounds. For completeness, Figure E1 traces the synthetic FVR bias in all three over the same country-mean sweep as Figure 1, at both latent dispersions. The Gini and CV divide by the mean; the below-5 share instead saturates toward 0 or 1 as the mean moves away from its fixed threshold. Different mechanisms, same outcome,

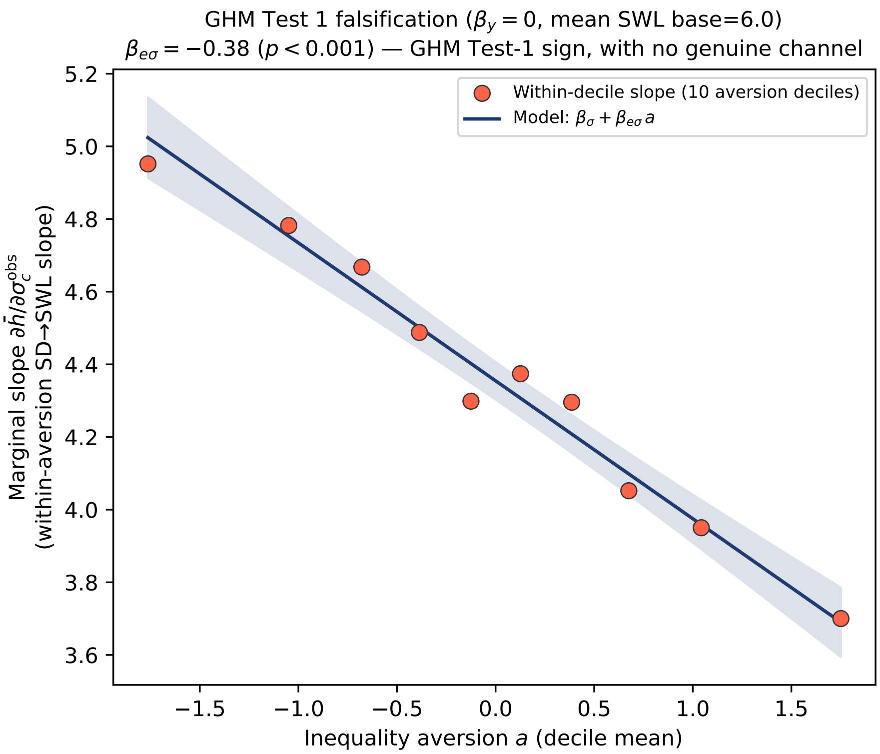

Figure D1: Synthetic falsification of GHM's Test 1 interaction. Goff et al.'s Test 1 (inequality-aversion interaction), evaluated in a synthetic cross-section built with *no* genuine wellbeing-dispersion channel and an inequality-aversion attitude that has *zero* effect on wellbeing by construction (Appendix C). Each point is one inequality-aversion decile; its height is the SD–SWL slope ($\partial\bar{h}/\partial\sigma_c^{\text{obs}}$) estimated *within* that decile. The slopes fall monotonically as professed aversion rises, and the slope of that decline is the interaction coefficient $\beta_{e\sigma} = -0.38$ ($p < 0.001$) — GHM's sign — even though the DGP contains no genuine dispersion channel, so FVR alone spuriously reproduces the test. (A flat line would mean no interaction; a coarse country-level median split, by contrast, shows near-equal subgroup slopes because the artefact is carried within countries by individual education, not by the between-country slope.)

visible in every panel below: a strong dependence on the mean that has nothing to do with any change in true dispersion.

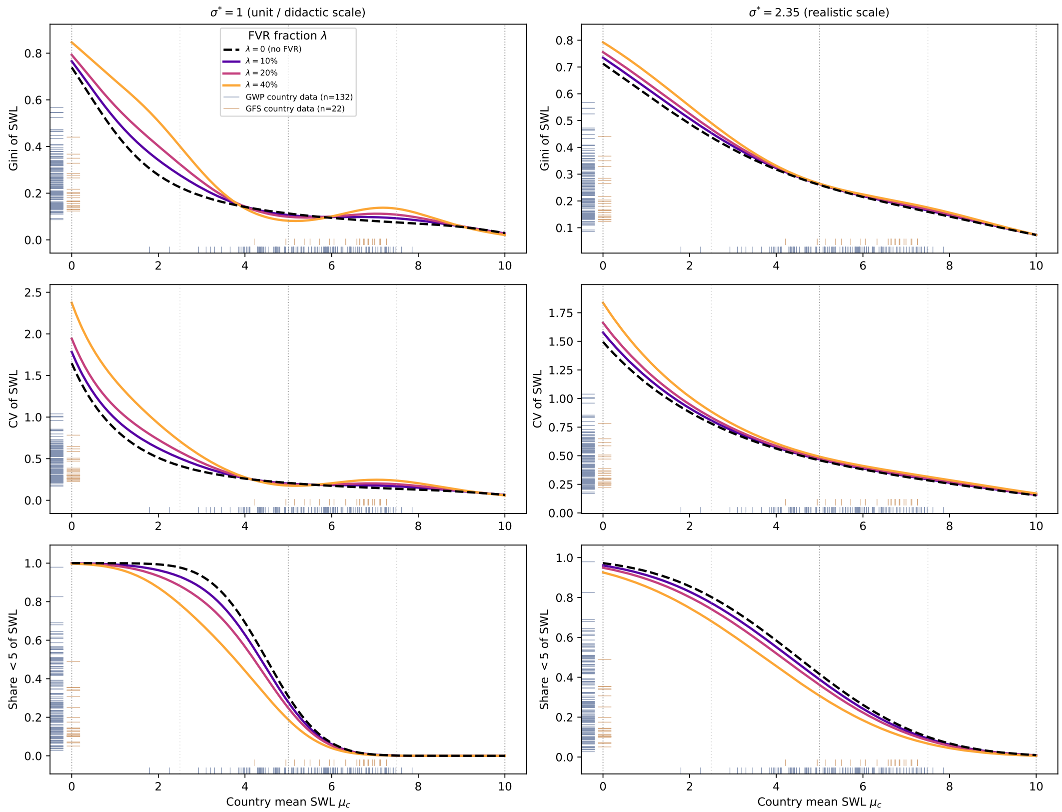

Figure E1: FVR **bias curves for the Gini coefficient, the CV, and the below-5 share.** FVR bias in the Gini coefficient, the coefficient of variation, and the share reporting below 5, as a function of the country mean μ, one curve per focal-responding share $\lambda \in \{0, 0.1, 0.2, 0.4\}$, at the unit scale $\sigma^* = 1$ (left) and the realistic scale $\sigma^* = 2.35$ (right); construction as in Figure 1, whose λ legend (top-left panel) applies here too. The Gini and CV divide by the mean; the below-5 share is instead evaluated at a fixed threshold — each carries a strong, mechanical dependence on the mean, unrelated to any change in true dispersion.

F Rare cases: latent distributions with strong endpoint mass

The correction leaves most countries' latent distributions with little mass at the scale endpoints. The exceptions are informative. Figure F1 shows the five Gallup Cantril-ladder countries — of the seven candidates with visibly heavy latent tails — with the largest latent endpoint mass $\hat{\pi}_c(0) + \hat{\pi}_c(10)$. These latent distributions are wide enough that ordered-probit censoring piles mass at 0 or 10 even after the FVR component is removed: qualitative evidence that something like censoring is at work in these data. Such cases are rare, which is also why the dispersion computed from the latent PMF stays close to the continuous latent scale σ_c for almost every country (Section 5.5).

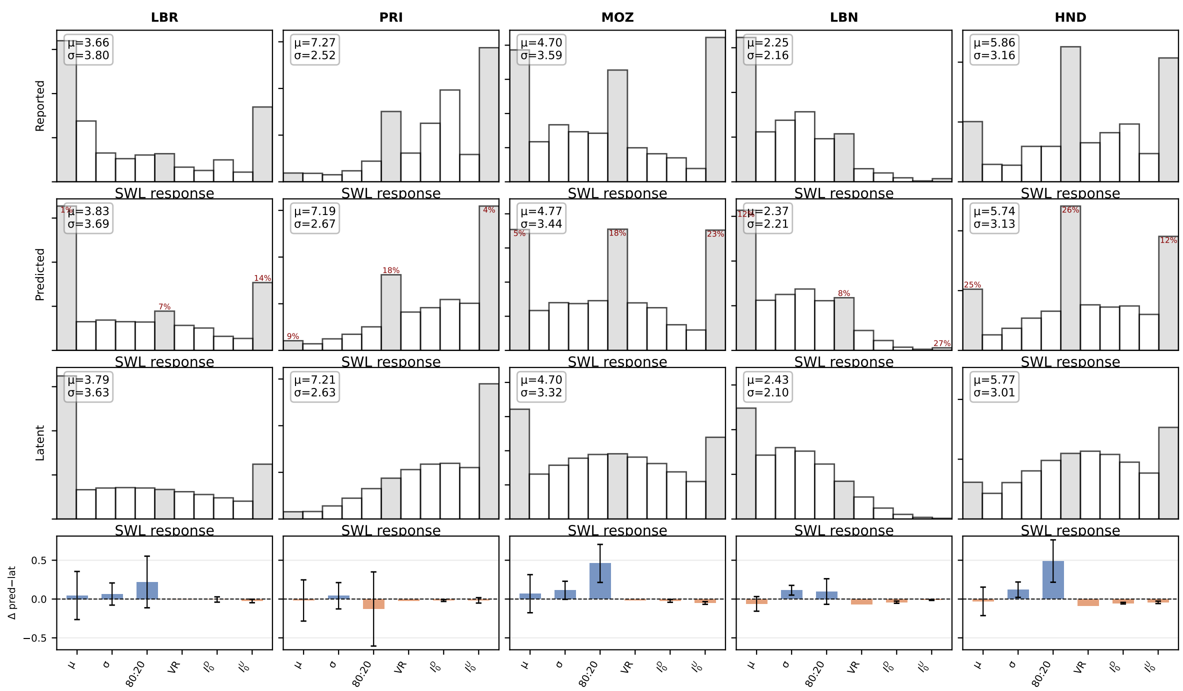

Figure F1: Rare cases where the latent distribution retains strong endpoint effects. The five Gallup Cantril-ladder countries with the largest latent endpoint mass $\hat{\pi}_c(0) + \hat{\pi}_c(10)$: Liberia (0.45), Puerto Rico (0.31), Mozambique (0.28), Lebanon (0.25), and Honduras (0.21). Rows show the reported, predicted, and latent PMFs, with the correction strip below, in the format of Figure 2. Even after the FVR component is removed, these latent distributions are wide enough that censoring piles mass at the scale endpoints; Puerto Rico is almost purely a top-end case and Lebanon almost purely bottom-end, while Liberia, Mozambique, and Honduras show mass at both ends.

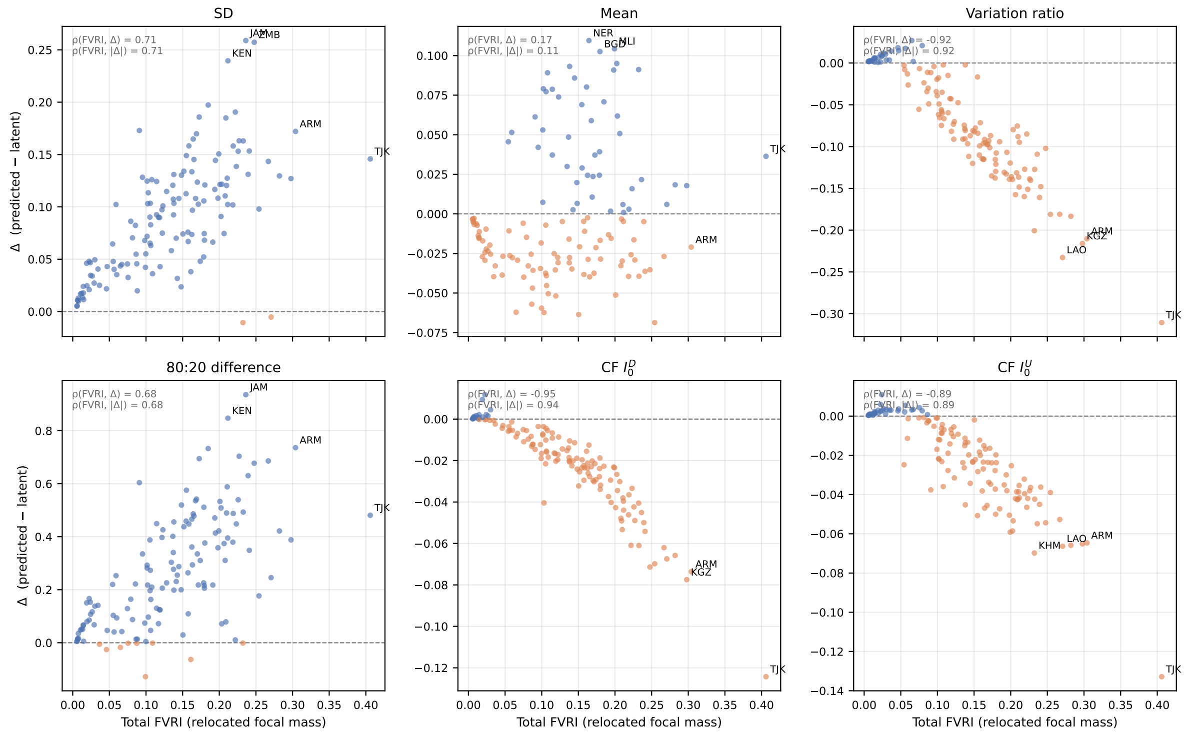

Figure G1: How much the FVR corrections cancel. Each panel plots, per country, the signed correction Δ = predicted − latent for one statistic against the country's overall FVRI (total relocated focal mass), Gallup Cantril ladder. For the dispersion and ordinal statistics the correction grows nearly monotonically with the relocated mass (Spearman correlations annotated), whereas for the mean a large relocated mass yields almost no net correction: rounding toward 10 and toward 0 largely offset. Extreme countries are labelled.

G How much the corrections cancel: total FVRI versus the net correction

Figure G1 plots, for each statistic, each country's net correction against its overall relocated focal mass. A statistic could in principle absorb a large relocated mass with no net change — the corrections at the three focal values can cancel — and the figure shows where that happens: for the mean, and essentially nowhere else.

H Replication on the Gallup satisfaction-with-life item

The Gallup World Poll fields a satisfaction-with-life item alongside the Cantril ladder, though in a different period and survey mode (2007–2010; N = 109,375, 114 countries). I fit the same reallocation model (specification (A)) and treat the result as an independent replication of the primary Cantril-ladder analysis rather than as an item comparison, since the period and mode differences confound any direct comparison of the two items.

Table H1: Scale of the FVR correction, Gallup satisfaction-with-life item. Scale of the FVR correction, Gallup satisfaction-with-life item (N = 109,375; 114 countries, 2007–2010). Columns as in Table 2.

	Pred range	Latent range	Median $\|\Delta\|$	Δ / range	Rank ρ
Mean	[2.409, 8.403]	[2.487, 8.423]	0.014	0%	1.000
SD	[1.074, 2.817]	[1.003, 2.743]	0.039	2%	0.984
80:20	[2.897, 8.019]	[2.709, 7.750]	0.081	2%	0.985
CF I_0^D	[0.622, 0.824]	[0.627, 0.829]	0.003	2%	0.897
CF I_0^U	[0.549, 0.818]	[0.550, 0.833]	0.002	1%	0.954
Variation ratio	[0.510, 0.858]	[0.615, 0.867]	0.011	3%	0.568

The primary results carry over. The standard-deviation correction is one-signed — FVR inflates the reported σ — in 103 of 114 countries (median signed $\Delta\sigma = +0.038$, largest +0.26 in Lao PDR), and the country ranking on the standard deviation survives the correction (Spearman $\rho = 0.984$, 96% credible interval $[0.970, 0.987]$). As in the main analysis the Cowell–Flachaire ordinal indices are more affected (I_0^D: $\rho = 0.897$; I_0^U: $\rho = 0.954$), though the effect on rankings is minimal. The GHM-style country regression (111 countries) gives raw/predicted/latent SD coefficients of -0.20/-0.19/-0.20 (latent CI $[-0.28, -0.06]$), so the cross-sectional association again survives the correction, with no significant attenuation.

I Replication on the Global Flourishing Study

The Global Flourishing Study's first wave asks both the Cantril "life today" item and the satisfaction-with-life item of the same respondents (22 countries; N = 177,353 and 177,157 respectively), although the survey question wording somewhat muddles the normal distinction (See Section 2 of Barrington-Leigh, 2026). This same-respondent design removes the period and question-wording confounds that muddy the comparison between the two Gallup datasets.

The primary results carry over. Corrections are one-signed in every GFS country (22 of 22 on the life-today item; 22 of 22 on the satisfaction item). SD rankings survive (GFS LT: $\rho = 0.998$, CI $[0.991, 0.999]$; GFS SWL: $\rho = 0.997$), levels move (GFS LT median $|\Delta\sigma| = 0.049$, largest +0.21 in Nigeria), and the Cowell–Flachaire indices are most affected (GFS LT I_0^D: $\rho = 0.784$, CI $[0.670, 0.828]$; I_0^U: $\rho = 0.718$); the rank instability is stronger in this 22-country sample than in the 135-country Gallup sample, perhaps simply because rank statistics in small samples are less stable. The GHM-style regression on GFS (22 countries) gives raw/predicted/latent SD coefficients of -1.00/-1.05/-0.96

Table I1: Scale of the FVR correction, GFS life-today item. Scale of the FVR correction, GFS life-today item ($N = 177{,}353$; 22 countries). Columns as in Table 2.

	Pred range	Latent range	Median $\lvert\Delta\rvert$	Δ / range	Rank ρ
Mean	[4.304, 7.366]	[4.239, 7.367]	0.014	0%	0.999
SD	[1.509, 3.531]	[1.505, 3.481]	0.049	2%	0.998
80:20	[4.176, 9.713]	[4.166, 9.591]	0.048	1%	0.994
CF I_0^D	[0.714, 0.819]	[0.741, 0.833]	0.005	5%	0.784
CF I_0^U	[0.688, 0.797]	[0.712, 0.824]	0.004	4%	0.718
Variation ratio	[0.689, 0.833]	[0.742, 0.871]	0.010	7%	0.036

Table I2: Scale of the FVR correction, GFS satisfaction-with-life item. Scale of the FVR correction, GFS satisfaction-with-life item ($N = 177{,}157$). Columns as in Table 2.

	Pred range	Latent range	Median $\lvert\Delta\rvert$	Δ / range	Rank ρ
Mean	[5.082, 8.029]	[5.085, 8.065]	0.010	0%	0.999
SD	[1.614, 3.720]	[1.611, 3.567]	0.050	2%	0.997
80:20	[4.412, 10.000]	[4.407, 9.777]	0.069	1%	1.000
CF I_0^D	[0.701, 0.822]	[0.757, 0.837]	0.003	2%	0.862
CF I_0^U	[0.529, 0.783]	[0.557, 0.787]	0.000	0%	0.879
Variation ratio	[0.561, 0.853]	[0.592, 0.880]	0.000	0%	0.910

(latent CI $[-1.08, -0.82]$). The satisfaction-with-life item of the same study behaves the same way: levels move (median $|\Delta\sigma| = 0.050$, largest +0.19 in Kenya), the Cowell–Flachaire indices are most affected (I_0^D: $\rho = 0.862$; I_0^U: $\rho = 0.879$), and the GHM-style regression (22 countries) gives raw/predicted/latent SD coefficients of -1.48/-1.59/-1.58 (latent CI $[-1.70, -1.43]$), so the cross-sectional association again survives the correction.

J Why the variation ratio is rank-unstable for life-today but not satisfaction-with-life

Section 4.4 reports that the variation ratio's predicted-to-latent rank correlation on the Gallup Cantril ladder is statistically indistinguishable from zero ($\rho = 0.016$, CI $[-0.035, 0.126]$). The same collapse recurs on the Global Flourishing Study's life-today item ($\rho = 0.036$, CI $[-0.076, 0.189]$; Table I1), yet the satisfaction-with-life item is far more rank-stable on both surveys — Gallup $\rho = 0.568$ (CI [0.532, 0.683]; Table H1), GFS $\rho = 0.910$ (CI [0.809, 0.945]; Table I2). Every other statistic in Table 2, Table H1, Table I1, and Table I2 retains a rank correlation of at least 0.7, so the variation ratio's life-today collapse is a genuine outlier, consistent across two independent surveys, and one that a defender of GHM's ordinal-measure argument might reasonably ask me to explain rather than merely report.

The mechanism is a discontinuity, not a magnitude. The variation ratio is $1 - \max_k p_k$: a function of only the single largest category share, so it is indifferent to how a correction reshapes the rest of a country's distribution but maximally sensitive to whichever shift, however small, changes *which* category is largest. The predicted and latent PMFs disagree about the modal category in 73 of 135 Gallup life-today countries and 8 of 22 GFS life-today countries, against only 35 of 114 Gallup satisfaction-with-life countries and 2 of 22 GFS satisfaction-with-life countries.

What separates the two items is how close each country's top two categories already are before any correction is applied. The margin between a country's largest and second-largest latent-PMF share has a median of only 0.009 on Gallup life-today (mean 0.018) and 0.010 on GFS life-today — a typical country's top two response categories separated by roughly one percentage point of the response mass — against 0.014 and 0.019 respectively for satisfaction-with-life. A wide margin survives a small correction; a razor-thin one does not. And the correction that closes those thin life-today margins is not generic: category 5, the central focal value already implicated in Section 2.2's below-5-share argument, is the modal category of the *predicted* (rounding-contaminated) distribution in 100 of 135 Gallup life-today countries (74%), even though the latent mode ranges over 8 distinct categories. FVR manufactures a spike at 5 that the de-rounded, latent distribution does not have; landing on a latent distribution that is itself often close to flat-topped across two or three adjacent categories, that manufactured spike is frequently enough to relocate the identified mode. Satisfaction-with-life responses are not similarly exposed: on GFS, for instance, category 10 — the top of the scale, not the centre — is the predicted mode in only 50% of countries, and the corresponding latent margin (0.019) is wide enough to survive the correction intact in nearly every case.

The instability is a property of the correction itself, not of the model's forward simulation. In every dataset, mismatches between the *raw, reported* modal category and the latent one are at least as large as predicted-versus-latent mismatches (Gallup life-today: 87 of 135 raw–latent mismatches against 25 raw–predicted mismatches; GFS life-today: 12 of 22 against 6). Predicted tracks raw closely, as the model is fit to reproduce the observed distribution; it is specifically the raw/predicted-to-latent step — the FVR correction — that relocates the mode. That the pattern recurs on both surveys, despite their different periods, modes, and question wordings, rules out a survey-specific artefact: what the two life-today items share is the item itself, a Cantril-ladder framing whose responses spread across the upper-middle of the 0–10 scale, straddling the central focal value, rather than clustering as decisively on a single category as satisfaction-with-life responses do.

None of this weakens the claim in Section 4.4: the Gallup life-today variation ratio's rank instability is real, replicates on an independent survey, and now has a concrete mechanism rather than only a synthetic analogue. But the mechanism is contingent on how decisively a population's modal response category is separated from its runner-up, not a generic property of ordinal, mode-based statistics under FVR. Being ordinal is no defence against FVR for the Gallup life-today item at the centre of GHM's claim; it should not, however, be read as evidence that every ordinal measure is this fragile for every item.

K Specification robustness and estimate quality

The primary estimates use specification (A): hierarchical per-country $\log \sigma_c$, country-varying β_S and β_N, cardinal cutpoints, and rounding-layer covariates comprising a constant and an education indicator. A companion specification (B), differing in the rounding-layer covariates and in a looser β_S prior, was run on all four datasets as a robustness check: per-country latent dispersion agrees across the two specifications at Spearman rank correlation 0.9998 and Pearson 0.9999 across all Gallup countries, with every discrepancy an order of magnitude inside the posterior standard error. Convergence: all chains pass divergence and posterior-predictive checks on all four datasets; diagnostics flag slow mixing (low effective sample size) for the per-country latent-scale parameters of a cluster of low-dispersion countries in the primary Gallup dataset, which affects the precision of those countries' interval statements but — per the cross-specification agreement above — not the paper's cross-country statistics.

Synthetic recovery of the hierarchical scale. The model was validated on a synthetic multi-group hierarchical per-country $\log \sigma_c$ reallocation data-generating process with a known truth. Figure K1 shows the two cases. When the true σ_c varies across groups (DGP A), the posterior recovers it closely (Pearson $r = 0.996$, RMSE $= 0.038$). When the true σ_c is constant (DGP B) — with the groups differing only in their FVR rates — the posterior does *not* manufacture spurious cross-country scale variation to absorb that heterogeneity: the posterior cross-country standard deviation of σ_c is 0.007 against a truth of zero. The model thus expresses genuine dispersion differences when they exist and reports their absence when they do not, which is the property the cross-country inequality claims rely on.

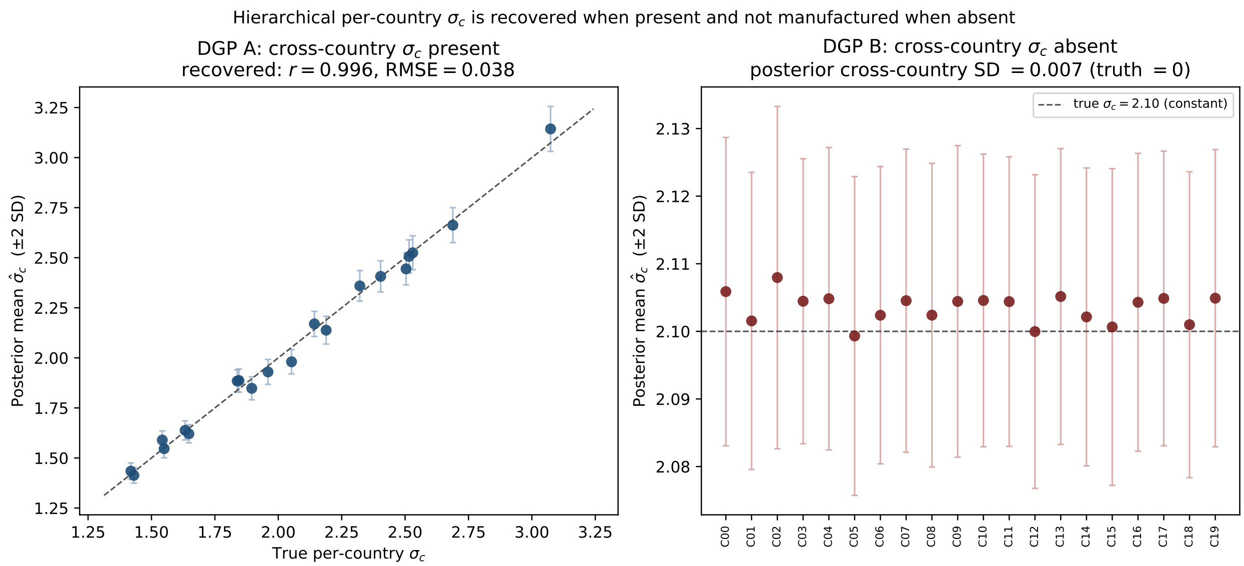

Figure K1: Synthetic recovery of the per-country scale σ_c. Synthetic recovery of the hierarchical per-country scale σ_c on a multi-group reallocation data-generating process. **(a)** DGP A, where the true σ_c varies across groups: posterior means (with ± 2 posterior SD) against the truth, hugging the identity line ($r = 0.996$, RMSE$= 0.038$). **(b)** DGP B, where the true σ_c is constant and groups differ only in FVR rate: the posterior places every group near the common truth, with cross-country posterior SD 0.007 (truth 0) — no spurious dispersion is manufactured.