Evenness under incomplete sampling: coupled error channels, hidden compensation, and the richness-estimation bottleneck

Luca Delvecchio
University of Salamanca, Salamanca, Spain
Email: lucadelvecchio975@gmail.com

September 2026

Abstract

1. Normalized evenness combines an abundance-diversity functional with richness, so incomplete sampling acts through two coupled errors. For $E_{\alpha,\beta} = L_\beta(D_\alpha)/L_\beta(S)$, the log-error decomposes exactly as $T_\beta = U_\beta + V_\beta$. With observed richness, V_β is nonnegative whenever species are missed, but total error can have either sign.

2. We generated 950,100 without-replacement rarefactions from 3,167 Ecological Register inventories and combined them with fixed-seed multinomial simulations of 250 known communities spanning five species-abundance families. At $\alpha = \beta = 1$ and 25% retention (median expected coverage 0.963), $\mathrm{Var}(U) + \mathrm{Var}(V) = 0.01887$, 4.17 times $\mathrm{Var}(T) = 0.00452$. The full normalized estimator was more rank-stable than either one-channel counterfactual (0.906 versus 0.804 and 0.851), while within-inventory Spearman channel dependence was exactly β-invariant. In known-truth simulations, channel dependence weakened with coverage in all 75 SAD$\times\alpha \times \beta$ combinations tested over coverage 0.80–0.95.

3. In Shannon/Pielou evenness, Chao–Wang–Jost entropy estimation materially improved joint correction. Chao1+CWJ had median raw-scale MSE ratios 0.96, 0.86, and 0.88 relative to the traditional estimator at coverages 0.80, 0.90, and 0.95; with true richness supplied as an oracle, the ratios were 0.49, 0.36, and 0.30. A paired log-MSE decomposition showed that the joint–oracle gap was mainly bias at coverage 0.80 but increasingly variance at 0.90 and 0.95. Richness error therefore remains, on median, the principal residual bottleneck, while negative covariance can conceal much of its marginal variance.

4. A coverage-standardized estimand changed the error geometry rather than merely shrinking error. The richness channel became two-sided, with negative values in about half of admissible replicates and median community mean near zero, while negative covariance persisted. Population evenness and evenness at standardized completeness should therefore be treated as distinct estimands. A supplementary proposition gives the eventual high-coverage decoupling limit under multinomial sampling.

Keywords: entropy estimation; evenness; Hill numbers; rarefaction; richness estimation; sample coverage; sampling error; Sharma–Mittal entropy.
Data/code for peer review: A reproducibility archive containing the analysis code and derived outputs is supplied with the submission for editors and reviewers.

1 Introduction

Evenness describes how abundance is distributed among the species of an assemblage. Since Pielou's normalized Shannon index, it has become a standard component of ecological diversity analysis (Pielou, 1966; Hill, 1973; Jost, 2010). In practice, however, evenness is calculated from a sample. Both the abundance distribution and the richness used to normalize it are therefore subject to incomplete observation.

Richness dependence has long been part of the discussion over evenness measures. Smith and Wilson (1996) treated independence from species richness as a central criterion, and Jost (2010) showed that the relation among diversity, richness, and evenness depends on the decomposition being used. The statistical issue considered here is more specific. We separate sampling error into the part inherited from effective diversity and the part introduced by richness normalization, determine the sign of the latter when observed richness is used, and examine the covariance between the two.

Sampling affects these components differently. Changes in observed relative abundances alter effective diversity; losses from the rare tail also reduce observed support, so $S_{\mathrm{obs}} < S$. Both changes enter the same normalized statistic. They can reinforce one another or they can offset. A stable evenness estimate does not by itself imply that either component is well estimated.

This issue is directly relevant to Alroy's (2025) critique of stand-alone evenness measures. In a large ecological corpus, deleting the dominant species changed rankings more strongly for Pielou's J than for the underlying Shannon effective number. That deterministic perturbation demonstrates sensitivity of the functional, but it does not identify the sampling contribution of effective diversity and richness separately. The analysis below addresses that narrower question for incomplete samples.

A related literature approaches incomplete observation by changing the comparison target. Chao and Jost (2012) developed coverage-based rarefaction and extrapolation, Chao et al. (2014) extended the framework to Hill diversity, and Chao et al. (2020) integrated completeness assessment, asymptotic diversity estimation, coverage-standardized comparison, and evenness. Their recommendation is especially relevant when true richness cannot be accurately inferred: evenness can instead be evaluated on a common, estimable assemblage fraction. Chisholm (2026) reaches a similar practical conclusion from a different starting point, recommending sample-level or coverage-based comparisons when community evenness is not reliably recoverable from an incomplete sample.

The population-target decomposition and coverage-standardized comparison answer different questions. The first keeps the population estimand fixed and exposes the two sources of error. The second changes the estimand to one defined at standardized completeness. We study both. This distinction matters because the samplewise inequality $S_{\mathrm{obs}} \leq S$ makes the population richness channel one-sided; no analogous deterministic order need hold between a coverage-standardized target $S(C)$ and its estimator $\widehat{S}(C)$.

We consider the normalized Sharma–Mittal evenness family (Sharma and Mittal, 1975)

$$E_{\alpha,\beta}(p) = \frac{L_\beta(D_\alpha(p))}{L_\beta(S)}. \tag{1}$$

Several canonical Hill-based evenness measures discussed by Chao and Ricotta (2019) arise as sections of this family; the representation question is not needed here. The present analysis requires only the definition above and the properties established in Section 2. The Hill order α determines abundance weighting in effective diversity. The second parameter β changes its cardinal representation relative to richness. Using several β sections makes it possible to ask whether sampling dependence changes, or whether only its numerical scale changes.

The empirical correction experiment concentrates on Shannon/Pielou evenness, $\alpha = \beta = 1$. Chao and Shen (2003) provide a coverage-adjusted estimator of Shannon entropy; Chao, Wang and Jost (2013) later derived a discovery-rate estimator that generally reduces entropy bias without requiring knowledge of total richness. The contrast is useful here because it improves the numerator without directly solving the denominator problem. Chao and Jost (2015) extend the discovery-rate approach to entropy and diversity profiles at general orders, but that broader estimation problem is outside the correction experiment studied below.

The paper examines five points: the exact decomposition of sampling error; the strength of dependence between its two components in ecological inventories; the effect of coverage and species-abundance structure on that dependence; the consequences of correcting numerator and denominator with alternative estimators; and the change in error geometry when the estimand itself is standardized by coverage. The analysis combines finite-population rarefaction of an Ecological Register corpus with fixed-seed known-truth multinomial simulations and a high-coverage asymptotic result.

2 Theory

2.1 Hill diversity and normalized evenness

Let $p = (p_1, \dots, p_S)$ be a probability vector on a finite positive support with $S \geq 2$, $p_i > 0$, and $\sum_i p_i = 1$. For $\alpha > 0$, $\alpha \neq 1$, Hill diversity is

$$D_\alpha(p) = \left(\sum_{i=1}^{S} p_i^\alpha \right)^{1/(1-\alpha)}, \tag{2}$$

with continuous Shannon limit

$$D_1(p) = \exp\left(-\sum_{i=1}^{S} p_i \log p_i \right). \tag{3}$$

For finite β define

$$L_\beta(x) = \begin{cases} \dfrac{x^{1-\beta} - 1}{1 - \beta}, & \beta \neq 1, \\ \log x, & \beta = 1. \end{cases} \tag{4}$$

For $x > 1$, $L_\beta(x) > 0$ and is strictly increasing. Equation (1) therefore lies in $[0, 1]$, equals one at complete uniformity, and changes monotonically with D_α at fixed (S, α). The sections $\beta = 0, 1, 2$ are respectively linear, logarithmic, and reciprocal/saturating cardinalizations.

2.2 Exact two-channel sampling decomposition

Suppose a sample yields empirical Hill diversity $\widehat{D}_\alpha$ and observed richness S_{obs}. Define

$$\widehat{E}_{\alpha,\beta} = \frac{L_\beta(\widehat{D}_\alpha)}{L_\beta(S_{\text{obs}})}. \tag{5}$$

Whenever $S_{\text{obs}} \geq 2$, the sample-to-population ratio factorizes exactly:

$$\frac{\widehat{E}_{\alpha,\beta}}{E_{\alpha,\beta}} = \underbrace{\frac{L_\beta(\widehat{D}_\alpha)}{L_\beta(D_\alpha)}}_{R_D} \underbrace{\frac{L_\beta(S)}{L_\beta(S_{\text{obs}})}}_{R_S}. \tag{6}$$

On the log scale set

$$U_\beta = \log R_D, \qquad V_\beta = \log R_S. \tag{7}$$

Then

$$T_\beta := \log \frac{\widehat{E}_{\alpha,\beta}}{E_{\alpha,\beta}} = U_\beta + V_\beta, \tag{8}$$

and

$$\operatorname{Var}(T_\beta) = \operatorname{Var}(U_\beta) + \operatorname{Var}(V_\beta) + 2\operatorname{Cov}(U_\beta, V_\beta). \tag{9}$$

Equation (8) is an algebraic identity; it defines the two error channels used throughout the paper.

Proposition 1 (One-sided richness-normalization channel)**.** *For every finite β and every sample with $2 \leq S_{\text{obs}} \leq S$,*

$$V_\beta = \log \frac{L_\beta(S)}{L_\beta(S_{\text{obs}})} \geq 0,$$

with equality if and only if $S_{\text{obs}} = S$.

Proof. For finite β, $L_\beta(x)$ is strictly increasing and positive for $x > 1$. Hence $S_{\text{obs}} \leq S$ implies $L_\beta(S_{\text{obs}}) \leq L_\beta(S)$, strictly when support is incomplete. □

The same monotonicity gives a useful conditional statement. If D_α were known and $S_L \leq S$ were only a lower bound for richness, then

$$\frac{L_\beta(D_\alpha)}{L_\beta(S_L)} \geq \frac{L_\beta(D_\alpha)}{L_\beta(S)}. \tag{10}$$

This is an upper-bound pressure from the richness channel, not an upper bound on a complete estimator in which D_α is also estimated. The numerator error can oppose it.

Proposition 2 (Exact within-inventory rank invariance across β)**.** *Fix an underlying inventory, Hill order $\alpha > 0$, and repeated samples for which $S_{\text{obs}} \geq 2$. For every finite β,*

$$\rho_S(U_\beta, V_\beta) = -\rho_S(\widehat{D}_\alpha, S_{\text{obs}}), \tag{11}$$

where ρ_S denotes Spearman rank correlation. The same sign-reversal statement holds for Kendall rank correlation whenever defined.

Proof. Let $g_\beta(x) = \log L_\beta(x)$, which is strictly increasing on $x > 1$. Within a fixed inventory, D_α and S are constants across repeated samples. Thus $U_\beta = g_\beta(\widehat{D}_\alpha) - g_\beta(D_\alpha)$ has the ranks of $\widehat{D}_\alpha$, whereas $V_\beta = g_\beta(S) - g_\beta(S_{\text{obs}})$ has the reverse ranks of S_{obs}. □

The proposition is deliberately within-inventory. After pooling different inventories, D_α and S are no longer common constants, so exact pooled rank invariance need not hold. At the boundary $\alpha = 0$, $D_0 = S$ and sample by sample $U_\beta = -V_\beta$, so $\widehat{E}_{0,\beta} = E_{0,\beta} = 1$. Near this boundary, numerical stability can be produced by cancellation rather than accurate estimation of either channel.

2.3 Coverage-standardized target

For a standardized assemblage fraction indexed by coverage C, define

$$E_{\alpha,\beta}(C) = \frac{L_\beta(D_\alpha(C))}{L_\beta(S(C))}. \tag{12}$$

Here $D_\alpha(C)$ and $S(C)$ are the diversity and richness evaluated at the same standardized coverage, following the rarefaction/extrapolation framework of Chao and Jost (2012), Chao et al. (2014), and Chao et al. (2020). If estimates $\widehat{D}_\alpha(C)$ and $\widehat{S}(C)$ are used, then

$$T_{\beta,C} = \log \frac{\widehat{E}_{\alpha,\beta}(C)}{E_{\alpha,\beta}(C)} = U_{\beta,C} + V_{\beta,C}, \tag{13}$$

with the same algebraic definitions as before. The decomposition survives the change of estimand. Proposition 1 does not: coverage standardization does not impose $\widehat{S}(C) \leq S(C)$ sample by sample.

2.4 Asymptotic limit as a supplementary result

The empirical analyses operate at moderate to high, but not rare-event, coverage. Appendix A proves that for a fixed finite full-support non-uniform community under multinomial sampling, channel correlation eventually approaches zero from below once support omission becomes exponentially rare and dominated by the least-abundant species. The limiting proposition is not used to explain the numerical magnitudes observed at coverage 0.80–0.95.

3 Materials and methods

3.1 Ecological Register inventories and relation to Alroy's sample

We used the public Ecological Register data associated with Alroy (2025), archived on Dryad (DOI 10.5061/dryad.brv15dvdc); Alroy's analysis code is archived separately on Zenodo (DOI 10.5281/zenodo.15762247). The raw register contained 82,870 taxon-by-inventory records representing 3,257 inventories. Following the archived code, missing values in `count` and `count.2` were set to zero and taxon abundance was defined as their sum. Inventories with fewer than three distinct positive abundance values were excluded, leaving 3,167.

Alroy's published paper reports an analytical database of 3,095 inventories, so the analytical samples are not identical. As a numerical cross-check, dominant-species removal on the 3,167-inventory corpus reproduced his rounded rank correlations. If an inventory contains species counts n_i and total abundance N, his Simpson-based diversity is

$$\widetilde{D}_{2,A} = \left[\sum_i \frac{n_i(n_i-1)}{N(N-1)}\right]^{-1}. \tag{14}$$

Spearman correlations between original and dominant-removed values were 0.9428 for Shannon effective diversity D_1, 0.8739 for $\widetilde{D}_{2,A}$, and 0.7551 for Pielou's J, matching Alroy's rounded 0.943, 0.874, and 0.755. Agreement of these rounded statistics does not imply identity of the analytical samples.

All sampling analyses below instead use the ordinary plug-in Hill functional

$$\widehat{D}_\alpha^{\text{plug}} = D_\alpha(\widehat{p}), \qquad \widehat{p}_i = n_i/N. \tag{15}$$

3.2 Finite-inventory rarefaction

Each complete Ecological Register inventory was treated as a finite empirical reference population, not as the true underlying ecological community. We rarefied without replacement using the observed taxon counts. Three sampling fractions were used: 25%, 50%, and 75% of the inventory's individuals. For each inventory and fraction, 100 independent multivariate hypergeometric samples were drawn with random seed 20260820, producing 950,100 rarefied samples in total.

For each replicate we calculated S_{obs}, plug-in Hill diversity, and the exact channels U_β and V_β for $\alpha \in \{0.10, 0.25, 0.50, 1, 2\}$ and $\beta \in \{0, 1, 2\}$. Replicates with $S_{\text{obs}} = 1$ were excluded because the logarithmic channel representation is undefined there; they comprised 0.3375% at 25% retention, 0.0284% at 50%, and 0.0016% at 75%.

Within each inventory we centered replicate values before pooling second moments, so reported variances and covariances represent within-inventory sampling variation rather than between-inventory heterogeneity. We also calculated each inventory's replicate correlation $\mathrm{Corr}(U, V)$ and summarized these by their median across inventories.

To connect the within-inventory decomposition to Alroy's cross-inventory rank criterion, each replicate number was treated as a complete rarefaction "world." For each retention fraction and replicate world, inventories were ranked by the full-inventory target and by four sampled quantities at $\alpha = 1$: unnormalized $\widehat{D}_1$; full normalized evenness; a diversity-error-only counterfactual $L_\beta(\widehat{D}_1)/L_\beta(S)$; and a richness-error-only counterfactual $L_\beta(D_1)/L_\beta(S_{\mathrm{obs}})$. Spearman correlations with the full-inventory target were summarized across the 100 worlds.

3.3 Richness and entropy correction

For richness we used the classical abundance-based Chao1 estimator (Chao, 1984). If f_1 and f_2 denote the numbers of singletons and doubletons,

$$\widehat{S}_{\mathrm{Chao1}} = \begin{cases} S_{\mathrm{obs}} + \dfrac{f_1^2}{2f_2}, & f_2 > 0, \\ S_{\mathrm{obs}} + \dfrac{f_1(f_1-1)}{2}, & f_2 = 0. \end{cases} \tag{16}$$

For Shannon entropy, the first corrected estimator was Chao–Shen (Chao and Shen, 2003). With sample size n, empirical probabilities $\widehat{p}_i = n_i/n$, estimated coverage $\widehat{C} = 1 - f_1/n$, and $\widetilde{p}_i = \widehat{C}\widehat{p}_i$,

$$\widehat{H}_{\mathrm{CS}} = -\sum_{i:n_i>0} \frac{\widetilde{p}_i \log \widetilde{p}_i}{1-(1-\widetilde{p}_i)^n}. \tag{17}$$

The second was the discovery-rate estimator of Chao, Wang and Jost (2013). Let ψ denote the digamma function and define

$$A = \begin{cases} \dfrac{2f_2}{(n-1)f_1 + 2f_2}, & f_2 > 0, \\ \dfrac{2}{(n-1)(f_1-1)+2}, & f_2 = 0,\ f_1 > 0, \\ 1, & f_1 = f_2 = 0. \end{cases} \tag{18}$$

Then

$$\begin{aligned} \widehat{H}_{\mathrm{CWJ}} = & \sum_{1 \le n_i \le n-1} \frac{n_i}{n}\left[\psi(n) - \psi(n_i)\right] \\ & + \frac{f_1}{n}(1-A)^{1-n}\left[-\log A - \sum_{r=1}^{n-1} \frac{(1-A)^r}{r}\right]. \end{aligned} \tag{19}$$

The implementation evaluates the tail term through the equivalent discovery-rate series to avoid cancellation in Eq. (19) when A is small. When $f_1 = f_2 = 0$ (so $A = 1$), the tail contribution is defined as zero and the closed form is not evaluated at its removable indeterminacy.

The known-truth experiment at $\alpha = \beta = 1$ used six main arms,

$$\{S_{\mathrm{obs}}, \widehat{S}_{\mathrm{Chao1}}\} \times \{\widehat{H}_{\mathrm{plug}}, \widehat{H}_{\mathrm{CS}}, \widehat{H}_{\mathrm{CWJ}}\}, \tag{20}$$

where evenness is $H/\log S$. This design compares denominator-only correction, two numerator corrections, and their joint versions against the traditional observed-richness plug-in statistic.

We also applied Chao1 to the without-replacement Ecological Register rarefactions as a diagnostic. Chao1 is not interpreted as a model-matched richness estimator for high-fraction finite-population sampling. That calculation is retained to show what can happen when an unseen-species correction is inserted into a sampling regime for which it was not designed.

3.4 Coverage calibration without changing the population target

A fixed fraction of individuals does not imply fixed sampling completeness. For an inventory with total count N, species count N_i, and a sample of size m drawn without replacement, finite-population expected coverage is

$$C(m) = 1 - \sum_{i=1}^{S} \frac{N_i}{N} \frac{\binom{N-N_i}{m}}{\binom{N}{m}}, \tag{21}$$

where the ratio is zero when $m > N - N_i$. For each inventory and target coverage we selected the smallest m satisfying Eq. (21). A dense grid $C \in \{0.80, 0.85, 0.90, 0.925, 0.95, 0.975, 0.99, 0.995\}$ was used for $\beta = 1$.

This is an oracle-completeness design: the complete finite inventory is known and is used to calibrate sampling effort, while the target remains full-inventory evenness. It therefore differs from the coverage-standardized estimand introduced below.

3.5 Known-truth SAD simulations

The finite-inventory experiment cannot distinguish sampling error relative to the observed inventory from sampling error relative to a generating community. We therefore generated 250 known communities, 50 from each of five SAD families. Richness was sampled uniformly from $S \in \{20, 40, 80, 120\}$. The families were:

(i) near-uniform: symmetric Dirichlet probabilities with concentration parameter drawn uniformly from 50 to 150;

(ii) lognormal: independent weights $w_i = \exp(Z_i)$ with $Z_i \sim N(0, \sigma^2)$ and $\sigma \sim U(0.7, 1.4)$;

(iii) log-series-like: truncated ranked weights $w_i = x^i/i$, $i = 1, \ldots, S$, with $x \sim U(0.94, 0.995)$;

(iv) geometric: ranked weights $w_i = r^{i-1}$ with $r \sim U(0.88, 0.97)$;

(v) dominant: one species assigned $p_{\max} \sim U(0.30, 0.70)$ and the remaining mass distributed according to normalized lognormal weights with $\sigma \sim U(0.5, 1.1)$.

Species labels were randomized after constructing ranked profiles. Each probability vector was known exactly.

Sampling was multinomial with fixed random seed 20260820. For each community and target $C \in \{0.80, 0.85, 0.90, 0.925, 0.95\}$, we selected the smallest integer n satisfying

$$C(n) = 1 - \sum_i p_i (1 - p_i)^n \geq C, \tag{22}$$

and generated 100 multinomial samples. The design comprised 125,000 samples with known S, p, D_α, and $E_{\alpha,\beta}$. The correction tables report coverages 0.80, 0.90, and 0.95 as anchor levels for compactness; the full five-level grid was used in the analyses.

For a community with replicate errors $T_r = \log(\widehat{E}_r/E)$, multiplicative log-bias is

$$B_{\log} = \exp\left(\frac{1}{R}\sum_{r=1}^{R} T_r\right) - 1. \tag{23}$$

Log-scale mean-squared error is

$$\mathrm{MSE}_{\log} = \frac{1}{R}\sum_r T_r^2 = \mathrm{Bias}(T)^2 + \mathrm{Var}(T). \tag{24}$$

3.6 Richness robustness and oracle benchmarks

Raw-scale MSE was computed directly. Joint correction was repeated with the improved Chao1 lower bound of Chiu et al. (2014), which uses tripletons and quadrupletons. For $f_4 > 0$ we used

$$\widehat{S}_{\text{iChao1}} = \widehat{S}_{\text{Chao1}} + \frac{n-3}{n}\frac{f_3}{4f_4}\max\left\{f_1 - \frac{n-3}{n-1}\frac{f_2 f_3}{2f_4}, 0\right\}, \tag{25}$$

with the correction set to zero when $f_4 = 0$. We also included oracle-richness benchmarks $\widehat{H}_{\text{CS}}/\log S$ and $\widehat{H}_{\text{CWJ}}/\log S$. These are diagnostic rather than implementable; they isolate the error remaining after denominator uncertainty is removed.

For each community and coverage, the Chao1+CWJ estimator was paired with its oracle-S counterpart on the same replicates. For every paired quantity we define Δ = joint minus oracle. The log-MSE difference was decomposed exactly as

$$\Delta\,\text{MSE}_{\log} = \Delta\,\text{Bias}(T)^2 + \Delta\,\text{Var}(T).$$

When $\Delta\,\text{MSE}_{\log} > 0$, we also recorded $\Delta\,\text{Var}(T)/\Delta\,\text{MSE}_{\log}$. This conditional diagnostic need not lie in $[0, 1]$: the variance difference can be negative even when total MSE is larger. It is used to distinguish a positive richness cost expressed mainly through systematic error from one expressed mainly through repeated-sampling variance.

3.7 Coverage-standardized estimand

The known-truth design also permits a direct comparison between population evenness and evenness at standardized completeness. For each generating community and target coverage C, the theoretical coverage curve in Eq. (22) was inverted to obtain the corresponding sample-size coordinate. The theoretical $q = 0$ and $q = 1$ Hill-number sampling curves of Chao et al. (2014) were evaluated at that same coordinate to define $S(C)$ and $D_1(C)$, and hence

$$E_C = \frac{\log D_1(C)}{\log S(C)}. \tag{26}$$

For each multinomial reference sample, the estimated coverage curve was inverted and the corresponding interpolation/extrapolation estimators of richness and Shannon diversity were evaluated at the same estimated coverage, yielding $\widehat{S}(C)$, $\widehat{D}_1(C)$ and $\widehat{E}_C$.

Following the short-range extrapolation rule used by Chao et al. (2020), a replicate target was retained only when the estimated sample size required to reach C did not exceed twice the reference sample size. The retained fractions were 0.960, 0.965, 0.974, 0.978, and 0.980 for target coverages 0.80, 0.85, 0.90, 0.925, and 0.95. This admissibility rule is directional: it removes reference samples whose estimated coverage curve would reach the target only beyond the short-range extrapolation horizon. Standardized results are therefore conditional on short-range estimability. The standardized channels were then computed from Eq. (13). This is a different estimand from population evenness and is not treated as a seventh correction arm.

3.8 Role of deterministic dominant removal

The dominant-removal analyses were retained as a structural benchmark and as a direct bridge to Alroy's perturbation experiment. They are not used as evidence for sampling covariance. Appendix B reports the exact change in $\log D_\alpha$ and the crossing order at which deleting the dominant species switches from reducing to increasing effective diversity.

4 Results

4.1 The two sampling channels are strongly negatively dependent

The rarefaction experiment produced strong negative within-inventory dependence. For $\beta = 1$, the median Pearson correlation at 25% retention was -0.9976 at $\alpha = 0.10$, -0.9837 at $\alpha = 0.25$, -0.9289 at $\alpha = 0.50$, -0.7281 at $\alpha = 1$, and -0.4380 at $\alpha = 2$ (Fig. 1). Near $\alpha = 0$, the exact identity $U = -V$ explains the almost perfect compensation.

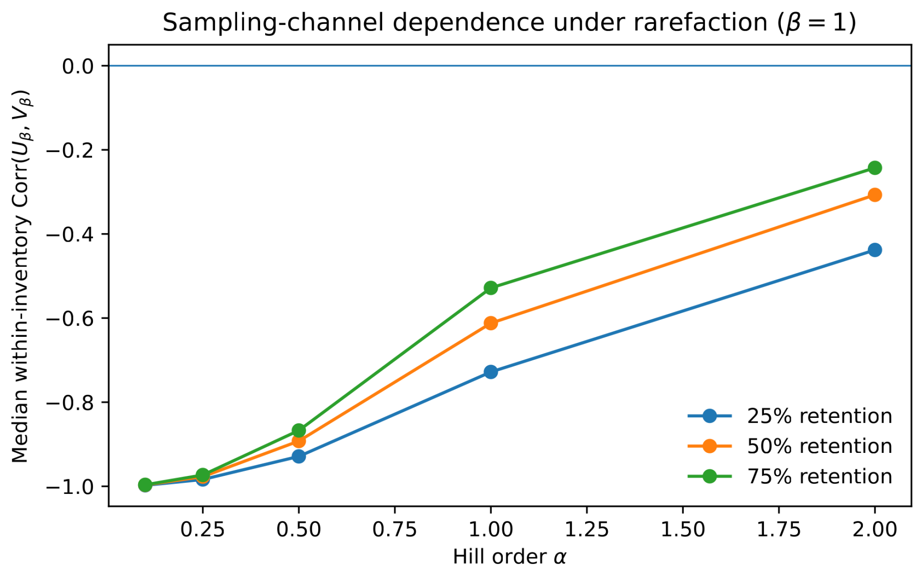

Figure 1: Within-inventory dependence of the two sampling channels in 3,167 Ecological Register inventories. Median replicate Pearson correlation between U_β and V_β under without-replacement rarefaction. The figure shows $\beta = 1$. Spearman dependence is exactly β-invariant by Proposition 2; Pearson dependence is not invariant, but its observed differences across $\beta = 0, 1, 2$ were small.

The covariance was quantitatively large. At $\alpha = \beta = 1$, 25% retention gave

$$\mathrm{Var}(U) = 0.011837, \quad \mathrm{Var}(V) = 0.007037, \quad 2\,\mathrm{Cov}(U, V) = -0.014352,$$

so that $\mathrm{Var}(U + V) = 0.004522$. The marginal variances therefore sum to 4.17 times the total variance. At 50% retention the corresponding values were 0.004855, 0.003330, -0.005688, and 0.002496; at 75%, 0.001685, 0.001444, -0.002004, and 0.001124 (Fig. 2).

4.2 Cross-inventory rank stability shows the same compensation in Alroy's metric

At 25% retention, the mean Spearman correlation with the full-inventory target was 0.9667 for unnormalized D_1, 0.9060 for full normalized evenness, 0.8043 when only the diversity channel was allowed to vary, and 0.8506 when only the richness channel was allowed to vary. At 50% retention the corresponding values were 0.9912, 0.9600, 0.9375, and 0.9379; at 75%, 0.9974, 0.9849, 0.9799, and 0.9778 (Fig. 3).

Unnormalized Hill diversity remains more rank-stable than normalized evenness under incomplete sampling, consistent with Alroy's empirical concern. At all three retention fractions, however, the full normalized estimator is more rank-stable than either one-channel-error counterfactual. The negative dependence between channels therefore preserves part of the cross-inventory ordering that either error would disrupt on its own.

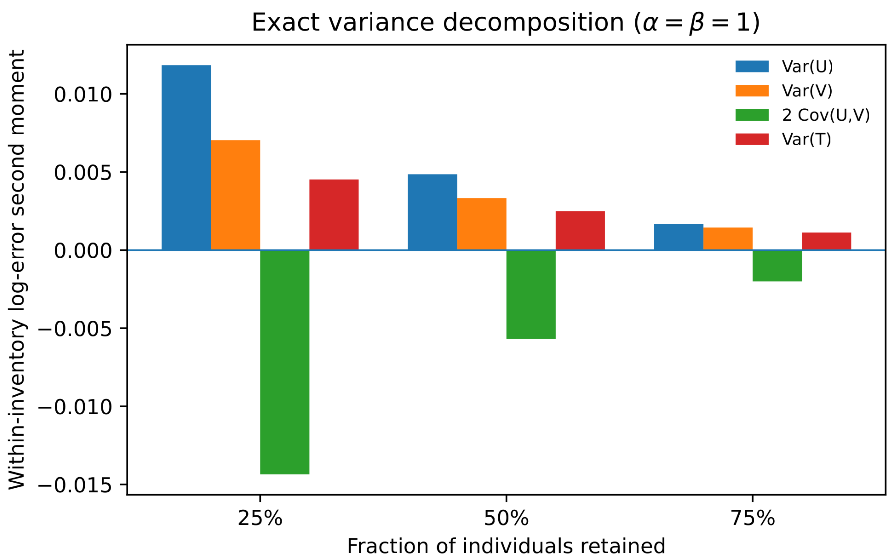

Figure 2: Exact within-inventory log-error variance decomposition at $\alpha = \beta = 1$. The negative covariance term is large relative to both marginal components.

4.3 The richness channel is one-sided, but total plug-in error is not

Proposition 1 fixes the sign of V only. In the Ecological Register rarefactions the net plug-in multiplicative log-bias happened to be positive. At $\alpha = 1$ and 25% retention, it was +28.03% for $\beta = 0$, +7.35% for $\beta = 1$, and +0.63% for $\beta = 2$; at 75% retention the corresponding values were +4.97%, +1.35%, and +0.10% (Fig. 4).

The known-truth simulations provide the important qualification. In the near-uniform family, plug-in Pielou evenness was negatively log-biased at every tested coverage: median multiplicative log-bias was −3.59% at coverage 0.80 and −3.38% at 0.95 even though $V \geq 0$ sample by sample. Hence

$$V \geq 0 \ \not\Rightarrow \ T \geq 0.$$

This family lies very close to the upper boundary of the index. Since $\widehat{D}_1 \leq S_{\mathrm{obs}}$, plug-in Pielou evenness satisfies $\widehat{E} \leq 1$ and therefore

$$T \leq -\log E.$$

Across the near-uniform generating communities, median true evenness was 0.99884. The corresponding median upper bound on positive multiplicative error, $E^{-1} - 1$, was only 0.12% (90th percentile 0.19%). The example is therefore asymmetric by construction: positive total error is strongly compressed near uniformity. It nevertheless makes the logical point of Proposition 1 sharper—the one-sided richness channel does not determine the sign of total error.

The ordering across β does not establish a statistical advantage for $\beta = 2$. For a small relative perturbation of D_α,

$$U_\beta \simeq a_\beta(D_\alpha)\, \Delta \log D_\alpha, \qquad a_\beta(D) = \frac{D^{1-\beta}}{L_\beta(D)}, \qquad a_1(D) = \frac{1}{\log D}.$$

The same relative error in effective diversity is therefore reported with a β-dependent elasticity. At large D, the reciprocal section $\beta = 2$ strongly compresses that error; smaller numerical motion is not evidence of intrinsically better estimation.

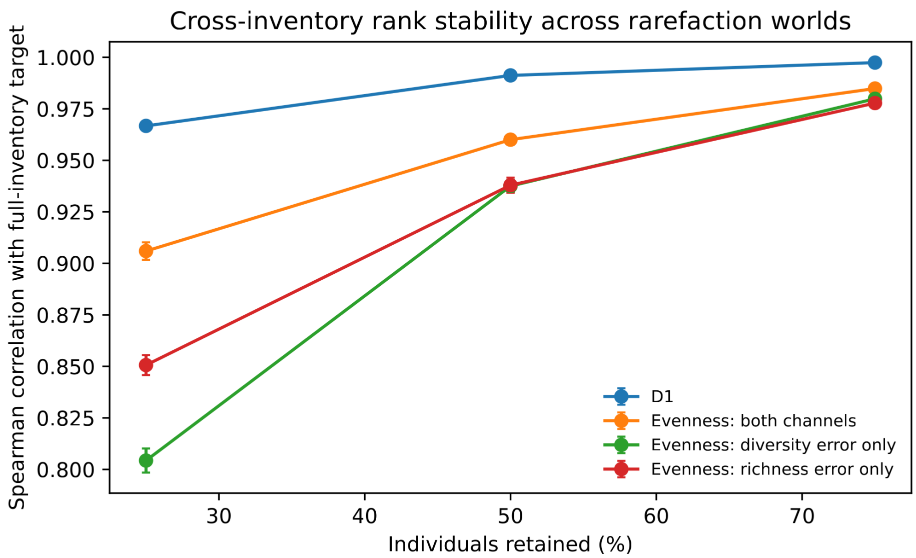

Figure 3: Direct bridge to Alroy's cross-inventory rank criterion. Mean Spearman correlation across 100 rarefaction worlds between the full-inventory target and the corresponding sampled quantity at $\alpha = \beta = 1$. Error bars show the standard deviation across worlds.

4.4 Chao1 in finite-population rarefaction remains a misspecification diagnostic

The Ecological Register finite-population rarefactions show what happens when Chao1 is used outside its model-matched setting. At $\alpha = \beta = 1$, median $\widehat{S}_{\text{Chao1}}/S$ was 0.933 at 25% retention, 1.045 at 50%, and 1.111 at 75%. The share of replicates with a negative Chao1 richness channel simultaneously rose from 0.34 to 0.54 and 0.67. At 75% retention the multiplicative log-bias changed from +1.35% with observed richness to −4.45% with Chao1, while total log-error variance rose from 0.00112 to 0.00451 (Table 1).

Table 1: Observed-richness versus Chao1 normalization in the Ecological Register rarefactions, $\alpha = \beta = 1$. Log-bias is the median community-level multiplicative log-bias. Variances are pooled within-inventory second moments after inventory-specific centering, whereas correlation is the median replicate correlation across inventories. log-RMSE is the median inventory-level square root of mean squared log-error.

Retention	Denom.	Log-bias (%)	Var(V)	Med. corr.	Var(T)	log-RMSE	Med. $\widehat{S}/S$
25%	S_{obs}	7.35	0.00704	-0.728	0.00452	0.084	–
25%	Chao1	-2.51	0.01980	-0.315	0.01373	0.117	0.933
50%	S_{obs}	3.39	0.00333	-0.612	0.00250	0.046	–
50%	Chao1	-3.85	0.01023	-0.221	0.00842	0.091	1.045
75%	S_{obs}	1.35	0.00144	-0.528	0.00112	0.024	–
75%	Chao1	-4.45	0.00513	-0.155	0.00451	0.074	1.111

4.5 The six-arm experiment separates numerator correction from richness correction

At coverage 0.80, median multiplicative log-bias was +6.58% for S_{obs} plus plug-in entropy, −7.69% for Chao1 plus plug-in entropy, +21.98% for S_{obs} plus Chao–Shen, and +5.30% for Chao1 plus Chao–Shen. The sign pattern shows bidirectional unmasking: correcting the denominator exposes

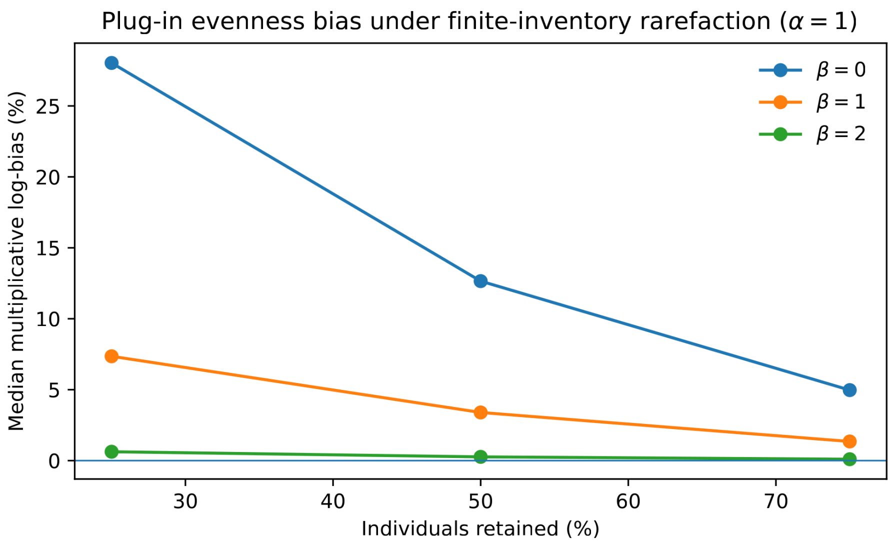

Figure 4: Median multiplicative log-bias under Ecological Register rarefaction at $\alpha = 1$. Positive net error in this corpus does not follow from Proposition 1 alone.

negative plug-in entropy error, whereas correcting the numerator while retaining observed richness exposes positive support error.

CWJ changes the size of the numerator correction. At coverage 0.80, S_{obs}+CWJ had +20.72% log-bias and Chao1+CWJ +4.41%; at coverage 0.95 the corresponding values were +6.24% and +0.77% (Table 2). The improvement is not simply a reduction in signed bias. Joint Chao1+CWJ also reduced median raw-scale MSE below the traditional estimator at all three reported coverages: 0.96, 0.86, and 0.88 times the traditional MSE at 0.80, 0.90, and 0.95. The corresponding median community-wise log-MSE ratios were 0.964, 0.883, and 0.916. The MSE advantage is therefore present on both scales in this design, although it is smaller on the log scale.

4.6 A second compensation appears after both components are corrected

The separate channel errors show why Chao–Shen and CWJ are not interchangeable in the near-uniform family. At coverage 0.80, median community mean errors were

$$U_{\mathrm{plug}} = -0.0987, \qquad U_{\mathrm{CS}} = +0.0135, \qquad U_{\mathrm{CWJ}} = +0.0006,$$

while

$$V_{\mathrm{obs}} = +0.0609, \qquad V_{\mathrm{Chao1}} = -0.0030.$$

These are separate cross-community medians of channel means and should not be added to reconstruct the median total error. Chao–Shen slightly overcorrects entropy in the same family in which Chao1 slightly overshoots richness; those corrected-channel errors partially cancel. CWJ removes most of the numerator overcorrection and the joint near-uniform log-bias changes from +0.94% with Chao1+Chao–Shen to −0.25% with Chao1+CWJ at coverage 0.80. The same mechanism remains smaller at coverage 0.95.

Across all communities the median numerator errors at coverage 0.80 were nearly identical for Chao–Shen and CWJ (−0.00845 and −0.00847), but this pooled similarity masks the family-specific sign difference. Figure 5 therefore reports channel means rather than only final evenness error.

Table 2: Six-arm correction experiment in the fixed-seed known-truth communities, $\alpha = \beta = 1$. Entries are medians across communities at three anchor coverages from the five-level simulation grid. Raw-MSE ratios are computed within each community relative to S_{obs}+plug-in before taking medians. Bias^2, variance, and RMSE are summarized separately, so the displayed medians need not satisfy the MSE identity even though it holds within every community. For the same reason, the marginal RMSE medians should not be used to infer the median paired log-MSE ratio between arms; the paired ratios are reported in the text.

Coverage	Estimator	Log-bias (%)	Bias^2	Variance	RMSE	Raw-MSE ratio
0.80	S_{obs}+plug-in	6.58	0.00406	0.00044	0.0767	1.00
0.80	Chao1+plug-in	-7.69	0.00791	0.00777	0.1287	1.75
0.80	S_{obs}+Chao–Shen	21.98	0.03948	0.00063	0.2015	5.31
0.80	S_{obs}+CWJ	20.72	0.03547	0.00106	0.1941	5.00
0.80	Chao1+Chao–Shen	5.30	0.00266	0.00680	0.1084	1.26
0.80	Chao1+CWJ	4.41	0.00186	0.00412	0.0871	0.96
0.90	S_{obs}+plug-in	3.71	0.00180	0.00039	0.0503	1.00
0.90	Chao1+plug-in	-4.68	0.00336	0.00371	0.0833	2.38
0.90	S_{obs}+Chao–Shen	12.39	0.01364	0.00027	0.1197	3.90
0.90	S_{obs}+CWJ	10.62	0.01019	0.00037	0.1050	3.29
0.90	Chao1+Chao–Shen	3.16	0.00097	0.00334	0.0804	1.22
0.90	Chao1+CWJ	1.83	0.00035	0.00264	0.0621	0.86
0.95	S_{obs}+plug-in	1.89	0.00095	0.00033	0.0407	1.00
0.95	Chao1+plug-in	-3.32	0.00164	0.00210	0.0628	2.08
0.95	S_{obs}+Chao–Shen	6.93	0.00449	0.00021	0.0705	2.58
0.95	S_{obs}+CWJ	6.24	0.00366	0.00026	0.0647	2.35
0.95	Chao1+Chao–Shen	1.41	0.00031	0.00189	0.0567	1.13
0.95	Chao1+CWJ	0.77	0.00013	0.00163	0.0468	0.88

4.7 The richness-estimation bottleneck changes character with coverage

Replacing Chao1 with iChao1 improved central richness recovery: median $\widehat{S}/S$ was 0.829 versus 0.806 for Chao1 at coverage 0.80, 0.933 versus 0.903 at 0.90, and 0.969 versus 0.948 at 0.95. With CWJ in the numerator, joint iChao1 raw-MSE ratios were 0.90, 0.82, and 0.87. The estimator is therefore not generically MSE-inferior once the numerator is updated.

The oracle comparison isolates the residual cost of estimating richness. With true richness and CWJ entropy, median raw-scale MSE fell to 0.49, 0.36, and 0.30 times the traditional estimator at coverages 0.80, 0.90, and 0.95; Chao1+CWJ gave 0.96, 0.86, and 0.88. For comparison, oracle-S+Chao–Shen gave 0.53, 0.44, and 0.34 at the same coverages (Fig. 6). True richness did not dominate community by community at lower coverage. At $C = 0.80$, Chao1+CWJ had lower log-MSE than oracle-S+CWJ in 28.8% of communities; the fraction fell to 3.6% at 0.90 and 1.2% at 0.95. On the raw scale the corresponding fractions were 26.0%, 3.6%, and 0.4%. A noisy, biased denominator can therefore improve the final statistic in individual communities when its error offsets the numerator error.

The variance pattern makes the same point from another direction. Using the separately displayed medians, total joint variance divided by oracle variance is 1.04, 1.74, and 2.25 at coverages 0.80, 0.90, and 0.95. At 0.80, negative covariance removes nearly all of the marginal denominator variance from the final estimator; as coverage rises, that cancellation weakens and denominator variance is transmitted more directly to total error.

Among communities with positive joint–oracle log-MSE gaps, the median variance share was 2.8%, 58.3%, and 82.6% at the three anchors, with IQRs $[-4.0, 37.3]$%, $[17.8, 95.0]$%, and $[36.1, 96.4]$%, respectively. This conditional summary is descriptive rather than an unconditional partition. The unconditioned paired medians lead to the same qualitative conclusion: median $\Delta\,\mathrm{Var}(T)$ was approximately -0.000046, $+0.000729$, and $+0.000708$, while median $\Delta\,\mathrm{MSE}_{\log}$ was $+0.001257$, $+0.001612$,

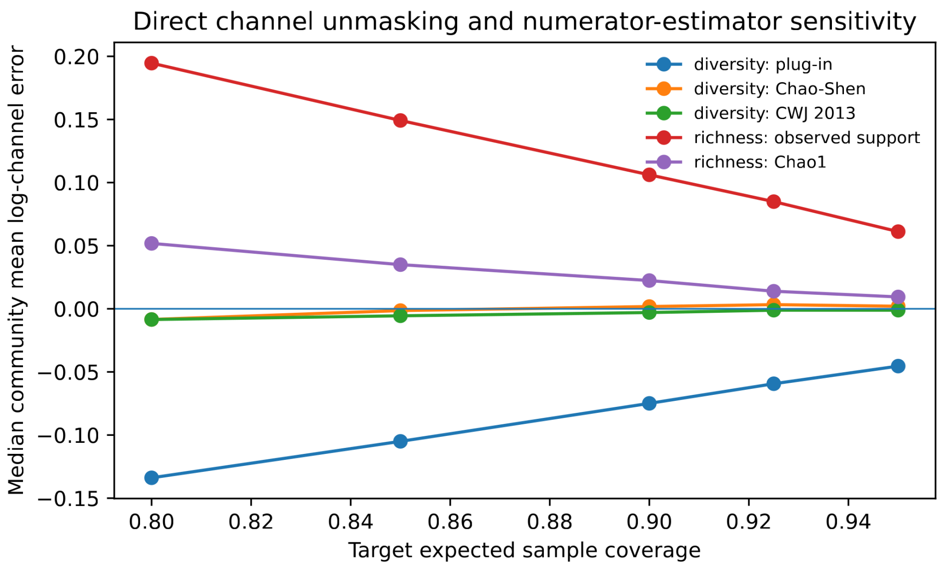

Figure 5: Direct channel unmasking in the fixed-seed known-truth simulation. Curves show medians across communities of replicate-mean log errors for plug-in, Chao–Shen, and CWJ diversity channels and for observed-support and Chao1 richness channels.

and +0.000995. These Δ summaries are medians of within-community paired differences and are not obtained by subtracting the separately summarized entries in Table 3. At $C = 0.80$ the median $\Delta\,\mathrm{Var}(T)$ is slightly negative: for the typical community, replacing Chao1 by true S increases variance marginally even while reducing log-MSE, so the oracle gain at this coverage comes through bias reduction rather than variance reduction. The residual richness cost is therefore mainly systematic at lower coverage and increasingly variance-driven at higher coverage in this design.

Table 3: Paired Chao1+CWJ and oracle-S+CWJ performance at three anchor coverages. Bias^2 and variance are medians of community-level log-error quantities. The two "raw ratio" columns are raw-scale MSE ratios computed within communities relative to S_{obs}+plug-in before taking medians; "Gap > 0" and "variance share" refer to log-MSE. Δ is joint minus oracle. For oracle-S+CWJ, $V \equiv 0$, so $\mathrm{Var}(T) = \mathrm{Var}(U_{\mathrm{CWJ}})$ by construction.

C	Joint Bias^2	Joint Var	Oracle Bias^2	Oracle Var	Joint raw ratio	Oracle raw ratio	Log gap > 0	Log var. share
0.80	0.001860	0.004118	0.000086	0.003968	0.958	0.486	0.712	0.028
0.90	0.000346	0.002636	0.000018	0.001518	0.860	0.357	0.964	0.583
0.95	0.000129	0.001626	0.000005	0.000724	0.876	0.303	0.988	0.826

The marginal channel variances tell a complementary story. For Chao1+CWJ, median $\mathrm{Var}(V_{\mathrm{Chao1}})$ exceeded median $\mathrm{Var}(U_{\mathrm{CWJ}})$ at every anchor coverage, while negative covariance kept much of that variance out of the final estimator (Table 4). The reported variance ratio is the median of community-wise ratios, not the ratio of the separately displayed medians. Large denominator variance can therefore be real at the component level without translating one-for-one into total estimator variance.

4.8 Coverage and abundance structure affect dependence differently

Fixed sampling fractions corresponded to heterogeneous expected coverage. At 25% retention, median finite-population expected coverage was 0.9628, with a 10th–90th percentile range 0.8118–0.9977. At 50%, median coverage was 0.9883 (0.9282–0.9995), and at 75% it was 0.9963 (0.9752–0.9999). A

Table 4: Corrected-channel variance decomposition for Chao1+CWJ. Entries are separate medians across communities; the exact identity $\mathrm{Var}(T) = \mathrm{Var}(U) + \mathrm{Var}(V) + 2\,\mathrm{Cov}(U, V)$ holds within each community, so the displayed medians need not close algebraically. The final column is the median of the community-wise ratio $\mathrm{Var}(V)/\,\mathrm{Var}(U)$.

C	$\mathrm{Var}(U_{\mathrm{CWJ}})$	$\mathrm{Var}(V_{\mathrm{Chao1}})$	2Cov	$\mathrm{Var}(T)$	Med. $\mathrm{Var}(V)/\mathrm{Var}(U)$
0.80	0.003968	0.006710	-0.006542	0.004118	1.89
0.90	0.001518	0.002824	-0.002097	0.002636	2.23
0.95	0.000724	0.001481	-0.000702	0.001626	2.32

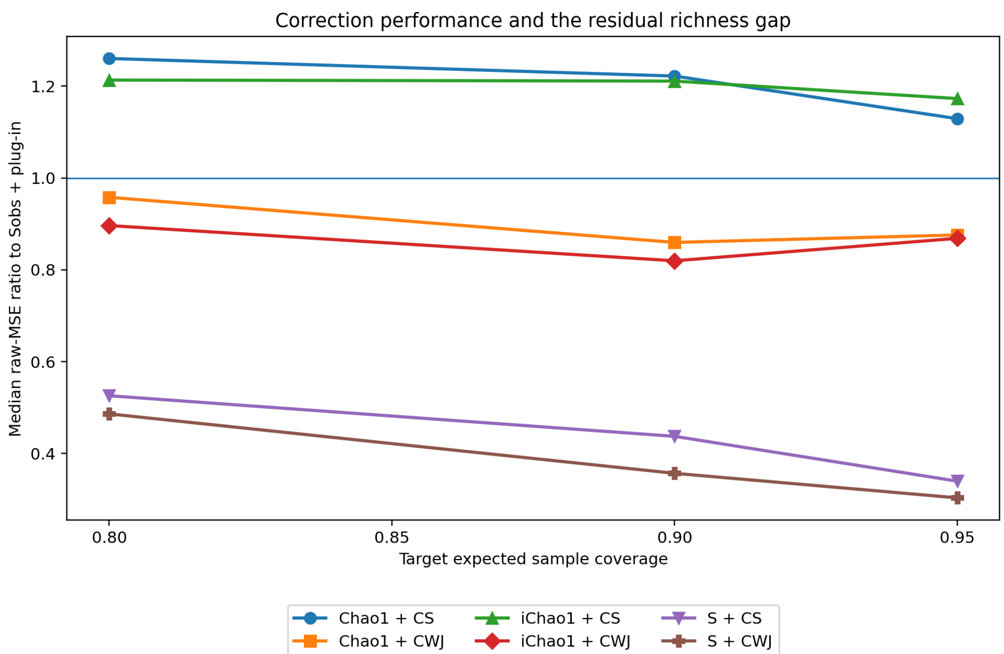

Figure 6: Raw-scale MSE robustness. Ratios are computed within each simulated community relative to the traditional S_{obs}+plug-in estimator and then summarized by the median. Oracle-S benchmarks remove denominator estimation error and are diagnostic rather than implementable.

common sampling fraction therefore represents different completeness levels across inventories.

In the known-truth design, $|\,\mathrm{Corr}(U, V)|$ decreased monotonically as target coverage rose from 0.80 to 0.95 in all 75 combinations of five SAD families, five α values, and three β values. At $\alpha = \beta = 1$, median absolute correlation changed from 0.814 to 0.544 for dominant communities, 0.919 to 0.742 for geometric communities, 0.866 to 0.650 for lognormal communities, 0.861 to 0.663 for log-series-like communities, and 0.923 to 0.796 for near-uniform communities (Fig. 7).

Relative channel variance was strongly SAD-dependent. Using the same orientation as Table 4, at $\alpha = \beta = 1$ median $\mathrm{Var}(V)/\,\mathrm{Var}(U)$ was roughly 0.62–0.77 in near-uniform and geometric communities and about 0.10–0.13 in dominant communities, with no common directional response to coverage. Coverage changes the coupling more consistently than it changes the allocation of variance between channels.

The oracle finite-inventory coverage-calibration design gave the same qualitative lesson: matching completeness reduced heterogeneity in channel dependence across assemblages, but did not erase abundance-structure effects. This design still targets the complete inventory and should not be confused with the standardized estimand in the next section.

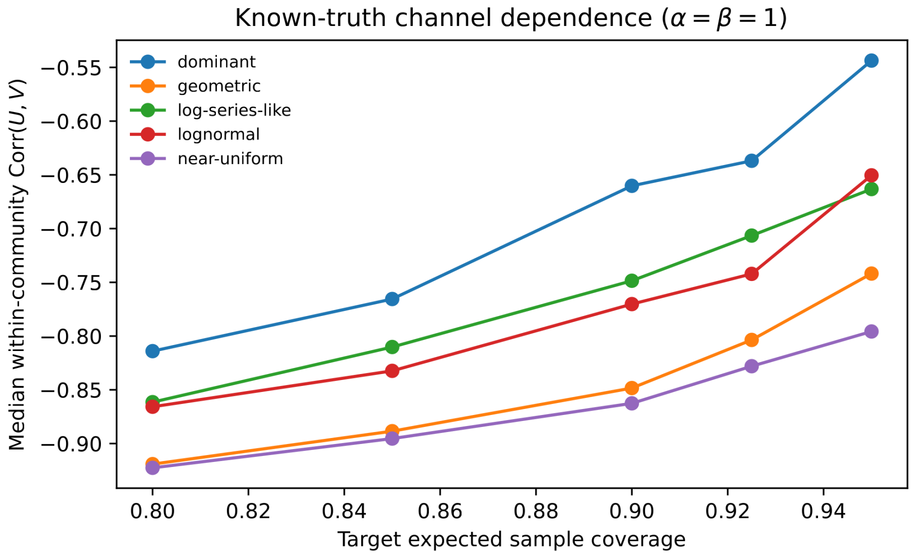

Figure 7: Known-truth simulations at $\alpha = \beta = 1$. Median within-community $\mathrm{Corr}(U, V)$ weakens with target expected coverage in every simulated SAD family. The same directional result holds in all 75 tested SAD$\times\alpha\times\beta$ combinations.

4.9 Coverage-standardized evenness changes the sign geometry of richness error

The population target and the coverage-standardized target behaved differently even when based on the same known communities. With population evenness, the observed-richness channel is nonnegative sample by sample. Median community mean V_{obs} was +0.199 at coverage 0.80, +0.109 at 0.90, and +0.062 at 0.95.

Under standardized completeness, the deterministic sign restriction disappeared. Among admissible replicates, the pooled share with $V_C < 0$ was 0.497 at coverage 0.80, 0.513 at 0.90, and 0.528 at 0.95. Median community mean V_C remained close to zero, while median multiplicative log-bias of $\widehat{E}_C$ stayed below 0.1% in magnitude over the full 0.80–0.95 grid (Table 5; Fig. 8).

Table 5: Coverage-standardized Shannon/Pielou evenness in admissible known-truth replicates at three anchor coverages. The extrapolation horizon is limited to twice the reference sample size. Variance and covariance entries are separate medians of within-community quantities and therefore need not close algebraically after summarization.

C	Retained	Med. log-bias (%)	Share $V_C < 0$	Med. mean V_C	$\mathrm{Var}(U_C)$	$\mathrm{Var}(V_C)$	2Cov	$\mathrm{Var}(T_C)$
0.80	0.960	-0.006	0.497	0.00034	0.00410	0.00358	-0.00688	0.00051
0.90	0.974	-0.063	0.513	-0.00120	0.00142	0.00146	-0.00239	0.00055
0.95	0.980	-0.099	0.528	-0.00137	0.00065	0.00066	-0.00086	0.00044

Negative dependence did not disappear. At coverage 0.80, the separate medians were $\mathrm{Var}(U_C) = 0.00410$, $\mathrm{Var}(V_C) = 0.00358$, $2\,\mathrm{Cov}(U_C, V_C) = -0.00688$, and $\mathrm{Var}(T_C) = 0.00051$. The exact variance identity holds within each community, not across these separately summarized medians. The apparent non-closure is largest at $C = 0.80$ because the cross-community component distributions are much wider there: the IQRs of $\mathrm{Var}(U_C)$, $\mathrm{Var}(V_C)$, and $2\,\mathrm{Cov}(U_C, V_C)$ are approximately 0.0115, 0.0102, and 0.0190 at 0.80, compared with 0.00123, 0.00143, and 0.00149 at 0.95. The IQR width of $\mathrm{Var}(T_C)$ likewise declines from 0.00264 at 0.80 to 0.00181 at 0.90 and 0.00112 at 0.95. Separately taken medians therefore need not correspond to the same part of the community distribution. As coverage rises, both marginal variances and the magnitude of the negative covariance shrink; the weakening of cancellation is why median $\mathrm{Var}(T_C)$ changes much less than either marginal variance. This pattern should not be read as a variance floor.

The admissibility filter also changes slightly with coverage: the retained fraction rises from 0.960 to 0.980 across the displayed rows. The selection is directional rather than random. Replicates excluded by the $2n$ horizon had mean standardized log-error T_C of about -0.063, -0.104, and -0.100 at coverages 0.80, 0.90, and 0.95, respectively, because their estimated coverage curves rose too slowly to reach the target within short-range extrapolation. Cross-coverage comparisons therefore use closely overlapping but not identical replicate sets, and the near-zero standardized biases reported here are conditional on admissibility. Coverage standardization removes the deterministic upward shift of the richness channel without removing stochastic coupling between the two channels. The median standardized log-bias drifts slightly downward across the displayed anchors (-0.006%, -0.063%, and -0.099%), but remains below 0.1% in magnitude; no substantive monotone bias claim is made from that small residual trend.

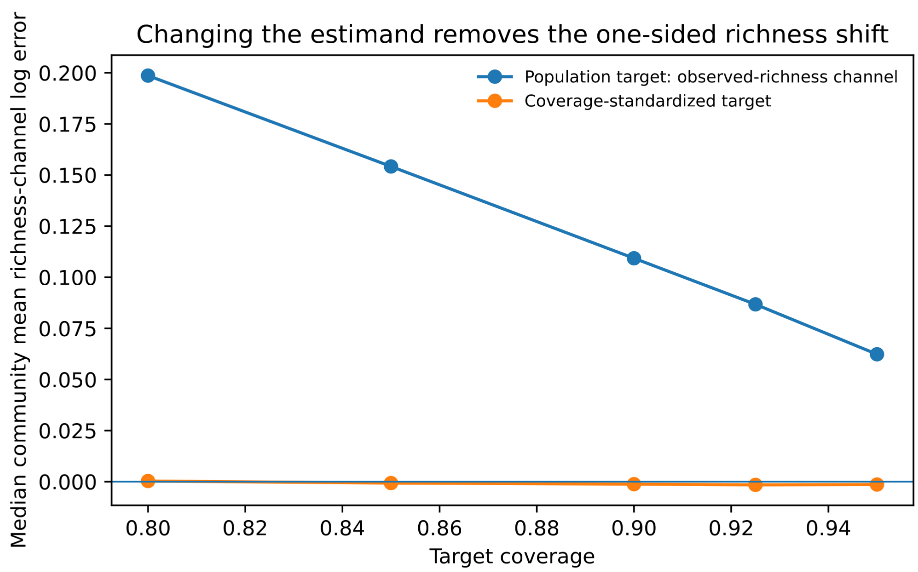

Figure 8: Richness-channel error under two estimands. For population evenness, observed support produces a positive richness-channel shift. At standardized completeness, the median community mean richness-channel error is near zero and its samplewise sign is two-sided.

4.10 Ordinal channel dependence is exactly β-invariant within inventories

Proposition 2 gives an exact ordinal result behind the empirical near-invariance. At fixed inventory and Hill order, Spearman dependence between U_β and V_β is identical for every finite β. Pearson correlation is not rank-invariant under nonlinear transformations and can vary with β. In the known-truth simulations, however, that variation was small: the median range of Pearson $\mathrm{Corr}(U, V)$ across $\beta = 0, 1, 2$ was 0.00004 at $\alpha = 0.10$, 0.00024 at 0.25, 0.00094 at 0.50, 0.0035 at 1, and 0.0073 at 2; the largest observed difference was 0.078. The small Pearson differences are consistent with curvature introduced by the cardinal transformation around an invariant rank structure. They provide no basis for a universal statistical default for β.

4.11 Operational coverage and the asymptotic appendix describe different regimes

The weakening observed over coverage 0.80–0.95 comes from finite-sample simulations, whereas Appendix A concerns the asymptotic regime. In the known-truth communities, the median expected number of unobserved species was about 24.5 at coverage 0.80 and 11.3 at coverage 0.95. At those coverages, support loss is still far from a one-missing-species regime. The proposition applies once omission is a rare event dominated by the least-abundant species. The simulations show that weaker coupling appears well before that limit; the agreement is directional.

Appendix A also reports an exact-enumeration check for $p = (0.6, 0.3, 0.1)$ and $\alpha = \beta = 1$. It verifies the asymptotic coefficient in a small community where the limiting regime can be reached. The moderate-coverage simulation curves are interpreted independently of this calculation.

5 Discussion

5.1 Coupled sampling error

Normalized evenness inherits sampling error through effective diversity and through richness. Equation (8) makes the distinction explicit on the log scale, where the two components add. With observed support, missing species always push the richness-normalization component upward. The diversity component has no fixed sign.

The covariance between them is large enough to change the interpretation of sampling stability. In the Ecological Register rarefactions at $\alpha = \beta = 1$ and 25% retention, $\mathrm{Var}(U) + \mathrm{Var}(V) = 0.01887$, 4.17 times the observed total variance $\mathrm{Var}(U + V) = 0.00452$. A small variance of the final index can therefore coexist with much larger variation in its two components because covariance cancels much of their contribution. The cancellation becomes nearly exact at low Hill orders, approaching $U = -V$ as $\alpha \to 0$.

5.2 Correction is estimator-specific, but the two-channel problem is not

Both entropy corrections show bidirectional unmasking: denominator correction exposes negative plug-in entropy error, while numerator correction with observed richness exposes the positive support channel. They differ materially in MSE. Chao1+CWJ outperforms the traditional estimator on median on both the raw scale (MSE ratios 0.958, 0.860, 0.876) and the log scale (0.964, 0.883, 0.916) at the three anchor coverages, whereas Chao1+Chao–Shen does not.

The more durable result is the separation between numerator and denominator uncertainty. CWJ improves the numerator without estimating total richness, which is precisely the setting in which the denominator can be isolated most cleanly. Once true S is supplied, raw MSE falls substantially below the implementable joint estimator. The paired decomposition shows that the residual richness cost changes character with coverage: at 0.80 it is mainly systematic error; the variance contribution increases with coverage and is dominant by 0.95. On median, richness is therefore the principal residual bottleneck here, with its bias and variance contributions changing across completeness levels.

The near-uniform family also shows why final bias alone is an unreliable diagnostic. Because true evenness is already close to one, positive plug-in error is strongly bounded by the scale itself. Within that restricted range, Chao–Shen overcorrects entropy slightly while Chao1 overshoots richness slightly; the two corrected errors partially cancel. CWJ removes most of the numerator overcorrection and changes the sign of the small residual joint bias. Hidden compensation can therefore occur not only between two naive components, but between two corrected components.

5.3 Coverage standardization changes the estimand and the error geometry

Chao et al. (2020) argue that when true richness cannot be accurately inferred, evenness of the entire assemblage should not be forced from the incomplete sample; comparisons can instead be made at a standardized assemblage fraction. The present decomposition shows a statistical consequence of that move. Under the population target, observed richness satisfies $S_{\text{obs}} \leq S$ and the richness channel is one-sided. Under the standardized target, the estimator fluctuates around a richness defined at the same completeness level and the deterministic sign restriction disappears.

The simulation result is stronger than a generic reduction in bias. V_C is genuinely two-sided, with approximately half the admissible replicates on either side of zero, while negative covariance between U_C and V_C remains substantial. Standardization changes the geometry of the inferential problem: it removes a structural asymmetry but not the dependence between errors. This also clarifies why coverage standardization should be treated as a distinct strategy rather than as another correction arm for population evenness.

5.4 Coverage and abundance structure

Increasing coverage weakened $|\operatorname{Corr}(U, V)|$ throughout the known-truth design. Species-abundance structure still shifted the level of the correlation at any fixed coverage and had a stronger effect on relative channel variance. Support recovery depends heavily on the rare tail, while effective diversity weights that tail according to α. Equal coverage standardizes represented probability mass, not the full shape of the SAD.

The finite-inventory and known-truth designs therefore agree only directionally: the former operated at median expected coverages 0.963–0.996 across the three retention fractions, whereas the known-truth simulations used 0.80–0.95. Across these different completeness ranges, matching completeness removes one source of heterogeneity but does not make sampling geometry universal. The coverage-standardized estimand goes further by redefining the target itself.

5.5 High-coverage limit

Appendix A describes a different regime from the main simulations. For a fixed finite non-uniform community under multinomial sampling, the two channels eventually decouple as completeness approaches one. At coverage 0.80–0.95, the simulated communities still omit many species, so the single least-abundant omission event that drives the asymptotic expansion is not yet dominant. The proposition supplies a boundary condition. The monotone weakening over the operational coverage range remains an empirical property of the simulated designs.

5.6 The role of β under incomplete sampling

Changing β cannot add abundance information when S and α are fixed; each section is a monotone transformation of the same D_α. Proposition 2 goes further for sampling error: within an inventory, Spearman dependence between U_β and V_β is identical for every finite β. Pearson dependence changed little across $\beta = 0, 1, 2$ in the simulations, even though the numerical size of the errors and their relative variances changed appreciably.

The sampling mechanism linking effective-diversity error to support loss is therefore largely stable across the tested β values, while the cardinal scale determines how strongly those errors appear in reported evenness. Small numerical bias for $\beta = 2$ in some settings partly reflects scale compression near one and should not be read as a universal statistical advantage.

5.7 Relation to Alroy and Chisholm

Alroy's results show that normalized evenness can be less stable than the underlying effective diversity under perturbation. The rarefaction results agree with that comparison: at 25% retention, mean cross-inventory Spearman correlation was 0.9667 for D_1 and 0.9060 for Pielou-type evenness. The decomposition adds a second point. The full normalized estimator was more rank-stable than either one-channel counterfactual because the two internal errors partly cancelled. Normalization introduces an additional source of sampling error and, at the same time, a covariance structure that can conceal part of it.

Chisholm (2026) notes that sample evenness often exceeds community evenness. Proposition 1 identifies a mechanism that favors that direction whenever observed richness is used, but the near-uniform simulations show why no universal sign follows. A negative numerator channel can outweigh the positive richness channel, while near the uniform boundary the index itself sharply limits positive total error. Coverage standardization avoids this deterministic support asymmetry by changing the estimand, consistent with the practical recommendation in Chisholm (2026) and Chao et al. (2020).

5.8 Practical implications

For empirical work, the sampling regime and treatment of richness should be stated explicitly. Comparisons across assemblages should be made at similar coverage when feasible, and population evenness should not be conflated with evenness at standardized completeness. If a population target is retained, correction of one component should be evaluated together with the remaining error in the other. When a final estimator has low RMSE, the channel decomposition can show whether that precision reflects small component errors or compensation between larger ones.

For severely incomplete communities, a coverage-standardized target may be preferable to an increasingly unstable attempt to recover total richness. That choice is substantive: it changes what is being estimated. The population and standardized quantities should therefore be labeled separately rather than treated as interchangeable estimates of one parameter.

5.9 Limitations

The empirical rarefactions treat each Ecological Register inventory as a finite reference population, while the known-truth simulations use multinomial sampling from specified SAD families. The asymptotic proposition is also multinomial and applies much closer to complete sampling than the operational simulations. Exactly uniform SADs are excluded from its first-order argument because the delta-method variance vanishes there.

The correction experiment is restricted to $\alpha = \beta = 1$. Chao1, iChao1, Chao–Shen, and CWJ provide informative test cases, but they do not exhaust the available estimators of richness or entropy. Chao and Jost (2015) provide a route to general-order discovery-rate estimation that is not developed here. The richness-estimation bottleneck documented here applies to the estimators and designs studied, not to richness estimation in general.

The coverage-standardized analysis is implemented in a known-truth setting so that the target curve is known. Real-data inference adds uncertainty in the estimated coverage curve and in extrapolation. We restrict standardized targets to the conventional short-range horizon of at most twice the reference sample size. This filter preferentially excludes reference samples with unusually slow estimated coverage accumulation; in the present simulation those excluded replicates had substantially more negative standardized error. The near-zero standardized bias reported here is therefore conditional on short-range admissibility and should not be extrapolated beyond that horizon. Finally, monotone weakening of channel dependence is established empirically over coverage 0.80–0.95 for the tested SADs and parameters; no global monotonicity theorem is claimed.

6 Conclusions

Incomplete sampling enters normalized evenness through two coupled error channels. With observed richness, the richness-normalization channel is one-sided, but total error is not. In the Ecological Register rarefactions at $\alpha = \beta = 1$ and 25% retention, the two marginal variances sum to 4.17 times the variance of the final log-error. Apparent stability can therefore be produced by cancellation.

The correction experiments show that this is not tied to one entropy estimator. Chao–Shen and CWJ produce different numerator errors, and the near-uniform family reveals compensation even between corrected components. CWJ materially improves joint correction: Chao1+CWJ beats the traditional estimator in median raw-scale MSE at all three reported coverages, yet oracle S improves it much further. The paired log-MSE decomposition shows that the residual richness cost is mainly bias at coverage 0.80 and increasingly variance at 0.90–0.95, while negative covariance hides part of the marginal denominator variance throughout. At lower coverage this compensation is strong enough that the implementable joint estimator can even outperform its oracle-S counterpart in some communities.

Changing the estimand produces a different result. At standardized completeness, the richness channel is no longer forced upward; its sign is approximately balanced across replicates and its mean is close to zero. Negative coupling with the diversity channel nevertheless persists. Coverage standardization removes a structural asymmetry without making the two errors independent.

These results support treating normalized evenness as a composite estimator. Sampling design, coverage, richness treatment, and estimand should be reported alongside the final index. Low apparent error can reflect compensation; successful correction depends on both components; and when total richness is information-limited, standardized completeness offers a different, more directly estimable target rather than a better estimate of the same one.

Data and code availability

The Ecological Register data are publicly archived at Dryad (DOI 10.5061/dryad.brv15dvdc), and Alroy's analysis code is archived at Zenodo (DOI 10.5281/zenodo.15762247). A reproducibility archive supplied with the submission contains the derived fixed-seed rarefaction summaries, the known-truth simulation code with Chao–Shen and CWJ estimators, the paired oracle-gap decomposition, the derived correction tables, and the coverage-standardized summaries used in this manuscript. The public raw archives are not redistributed. The archive is supplied for peer review; the analysis code in the archive is released under the MIT License. On acceptance the archive will be deposited in a public repository and assigned a persistent identifier.

Conflict of interest

The author declares no conflict of interest.

A High-coverage decoupling: proposition and proof

Consider multinomial sampling of size n from a fixed finite community p. Define

$$F_{\alpha,\beta}(p) = \log L_\beta(D_\alpha(p)), \tag{27}$$

$$U_{\beta,n} = F_{\alpha,\beta}(\widehat{p}_n) - F_{\alpha,\beta}(p), \qquad V_{\beta,n} = \log L_\beta(S) - \log L_\beta(S_{\mathrm{obs},n}). \tag{28}$$

Let

$$p_* = \min_i p_i, \qquad m_* = \#\{i : p_i = p_*\},$$

and for a least-abundant species let

$$\Delta F_* = F_{\alpha,\beta}(p^{(-*)}) - F_{\alpha,\beta}(p) < 0,$$

where $p^{(-*)}$ is the distribution obtained by deleting that species and renormalizing the remainder.

Proposition 3 (High-coverage decoupling)**.** *Let $S \geq 3$. Suppose p has full support and is non-uniform, and fix $\alpha > 0$ and finite β. Under multinomial sampling, compute channel correlation conditional on $S_{\mathrm{obs},n} \geq 2$. Then*

$$\mathrm{Corr}(U_{\beta,n}, V_{\beta,n}) \sim \frac{\Delta F_*}{\sigma_{\alpha,\beta}} \sqrt{m_* n}\,(1-p_*)^{n/2}. \tag{29}$$

Here $\sigma^2_{\alpha,\beta}$ is the nonzero first-order asymptotic variance coefficient of $F_{\alpha,\beta}(\widehat{p}_n)$. Consequently,

$$\mathrm{Corr}(U_{\beta,n}, V_{\beta,n}) \to 0^-,$$

and the absolute correlation is eventually strictly decreasing. True expected multinomial sample coverage satisfies

$$C_n = 1 - \sum_i p_i(1-p_i)^n, \tag{30}$$

and

$$|\,\mathrm{Corr}(U_{\beta,n}, V_{\beta,n})| = \Theta\left(\sqrt{(1-C_n)\log\frac{1}{1-C_n}}\right). \tag{31}$$

Proof. All moments are conditional on $A_n = \{S_{\mathrm{obs},n} \geq 2\}$. For $S \geq 3$, the excluded event has probability $o((1-p_*)^n)$ and does not affect leading orders. Let M_i be the event that species i is absent. For a least-abundant species, $P(M_i) = (1-p_*)^n =: q_n$. If $p_j > p_*$, then $P(M_j)/q_n \to 0$, and for distinct i, j, $P(M_i \cap M_j) = o(q_n)$. Thus exactly one least-abundant species is omitted with probability $m_* q_n[1 + o(1)]$.

Conditional on that omission, the empirical composition of the remaining species converges to $p^{(-*)}$. This vector strictly majorizes p, so strict Schur concavity of Hill diversity implies $D_\alpha(p^{(-*)}) < D_\alpha(p)$ for $\alpha > 0$. Since $x \mapsto \log L_\beta(x)$ is increasing, $\Delta F_* < 0$.

Let $v_\beta = \log L_\beta(S) - \log L_\beta(S-1) > 0$. On the dominant support-omission event, $V_{\beta,n} = v_\beta$ and $U_{\beta,n} \to \Delta F_*$. Therefore

$$\mathrm{Var}(V_{\beta,n}) = m_* q_n v_\beta^2 + o(q_n),$$
$$\mathrm{Cov}(U_{\beta,n}, V_{\beta,n}) = m_* q_n \Delta F_* v_\beta + o(q_n).$$

For the diversity channel, $F_{\alpha,\beta}$ is smooth at an interior non-uniform p. The influence function of $\log D_\alpha$ is

$$\xi_\alpha(i) = \begin{cases} \dfrac{\alpha}{1-\alpha}\left[\dfrac{p_i^{\alpha-1}}{\sum_j p_j^\alpha} - 1\right], & \alpha \neq 1, \\ -\log p_i - H(p), & \alpha = 1. \end{cases}$$

Writing

$$a_\beta(D) = \frac{D^{1-\beta}}{L_\beta(D)}, \qquad a_1(D) = \frac{1}{\log D},$$

the influence function of $F_{\alpha,\beta}$ is $\psi_{\alpha,\beta}(i) = a_\beta(D_\alpha)\xi_\alpha(i)$, with positive variance $\sigma^2_{\alpha,\beta} = \sum_i p_i \psi_{\alpha,\beta}(i)^2$ for a non-uniform community. The multivariate delta method gives

$$\mathrm{Var}(U_{\beta,n}) = \frac{\sigma^2_{\alpha,\beta}}{n} + o(n^{-1}).$$

Combining the expansions yields Eq. (29). Eventual strict decrease follows because

$$\frac{\sqrt{n+1}(1-p_*)^{(n+1)/2}}{\sqrt{n}(1-p_*)^{n/2}} \to \sqrt{1-p_*} < 1.$$

Finally, $1 - C_n \sim m_* p_* (1-p_*)^n$, which gives Eq. (31). $\square$

Exact-enumeration check

The limiting coefficient was checked independently by exact enumeration for $p = (0.6, 0.3, 0.1)$ and $\alpha = \beta = 1$.

n	Exact correlation	Asymptotic approximation	Exact/approx.
50	-0.247452	-0.279328	0.8859
80	-0.067602	-0.072747	0.9293
120	-0.010325	-0.010832	0.9532
180	-0.000545	-0.000562	0.9690

The ratio approaches one as both quantities tend to zero.

B Deterministic dominant-species removal

If the dominant species has relative abundance p_m, then for $\alpha \neq 1$ deleting it and renormalizing the remaining species gives

$$\Delta \log D_\alpha = \frac{1}{1-\alpha} \log \left(1 - \frac{p_m^\alpha}{\sum_i p_i^\alpha} \right) - \frac{\alpha}{1-\alpha} \log(1 - p_m). \tag{32}$$

Across the 3,167 Ecological Register inventories, using the plug-in Hill functional, the median net change in $\log D_\alpha$ changed sign with order: approximately -0.040 at $\alpha = 0.10$, -0.018 at 0.25, $+0.009$ at 0.50, $+0.042$ at 1, $+0.068$ at 2, and $+0.081$ at 4. A crossing order α_c satisfying $D_\alpha^- = D_\alpha$ was found within $[0.02, 4]$ for 72.1% of inventories. Its median was 0.256 (10th–90th percentile 0.074–0.979) and it was negatively associated with richness (Spearman $\rho \approx -0.465$). This median crossing order need not coincide with the order at which the median $\Delta \log D_\alpha$ over all inventories changes sign, because α_c is defined only for the 72.1% of inventories with a crossing in the search interval and the two summaries take medians over different objects. These are structural perturbation results, not sampling-covariance results.

C Sampling-balance diagnostic

For diagnostic purposes define β_{bal} by

$$\mathrm{Cov}(U_\beta + V_\beta, V_\beta) = 0, \qquad \text{equivalently} \qquad \frac{\mathrm{Cov}(U_\beta, V_\beta)}{\mathrm{Var}(V_\beta)} = -1.$$

In the Ecological Register fixed-fraction rarefactions a unique root over $\beta \in [0, 4]$ was found for every tested $(\alpha, \text{fraction})$ combination. For $\alpha = 1$, the root shifted from 0.949 at 25% retention to 1.657 at 50% and 2.416 at 75%. This is not an optimization criterion: both channels change with β, so the root balances simultaneously reparameterized errors and becomes poorly conditioned near $\alpha = 0$ and near complete sampling.

D Family-specific six-arm log-bias

Table 6 shows the family structure behind the pooled medians. The near-uniform family is the only one with negative traditional bias at the displayed coverages. CWJ almost removes the Chao–Shen overcorrection in that family, but its signed-bias advantage is SAD-specific: at $C = 0.80$ Chao1+CWJ has smaller absolute joint bias than Chao1+Chao–Shen in two of five families, and at $C = 0.95$ in three of five. This heterogeneity does not extend to overall error in the same way: median community-level log-MSE is lower for Chao1+CWJ than for Chao1+Chao–Shen in all five families at both displayed coverages.

Table 6: Median multiplicative log-bias (%) by SAD family in the fixed-seed six-arm simulation.

Family	C	S_{obs}+plug	Chao1+plug	S_{obs}+CS	S_{obs}+CWJ	Chao1+CS	Chao1+CWJ
Dominant	.80	8.94	-10.22	30.99	25.73	8.27	3.57
Dominant	.95	1.98	-3.22	9.74	6.79	3.68	1.00
Geometric	.80	11.55	-2.91	24.24	25.92	8.86	10.38
Geometric	.95	6.65	-1.00	10.68	11.41	2.57	3.14
Lognormal	.80	5.62	-5.52	16.54	17.31	4.46	4.76
Lognormal	.95	2.21	-2.25	5.91	6.14	1.26	1.10
Log-series-like	.80	21.12	-1.24	36.25	38.13	12.13	13.23
Log-series-like	.95	7.69	-1.08	13.01	11.78	3.84	2.74
Near-uniform	.80	-3.59	-9.52	7.67	6.50	0.94	-0.25
Near-uniform	.95	-3.38	-4.91	1.76	1.39	0.21	-0.23

References

[1] Alroy, J. (2025). Does evenness even exist? *Ecology Letters*, 28(7), e70181. doi:10.1111/ele.70181.

[2] Chao, A. (1984). Nonparametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics*, 11, 265–270.

[3] Chao, A., and Jost, L. (2012). Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. *Ecology*, 93, 2533–2547. doi:10.1890/11-1952.1.

[4] Chao, A., Gotelli, N. J., Hsieh, T. C., Sander, E. L., Ma, K. H., Colwell, R. K., and Ellison, A. M. (2014). Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecological Monographs*, 84, 45–67. doi:10.1890/13-0133.1.

[5] Chao, A., and Ricotta, C. (2019). Quantifying evenness and linking it to diversity, beta diversity, and similarity. *Ecology*, 100, e02852. doi:10.1002/ecy.2852.

[6] Chao, A., and Shen, T.-J. (2003). Nonparametric estimation of Shannon's index of diversity when there are unseen species in sample. *Environmental and Ecological Statistics*, 10, 429–443. doi:10.1023/A:1026096204727.

[7] Chao, A., Wang, Y.-T., and Jost, L. (2013). Entropy and the species accumulation curve: a novel entropy estimator via discovery rates of new species. *Methods in Ecology and Evolution*, 4, 1091–1100. doi:10.1111/2041-210X.12108.

[8] Chao, A., and Jost, L. (2015). Estimating diversity and entropy profiles via discovery rates of new species. *Methods in Ecology and Evolution*, 6, 873–882. doi:10.1111/2041-210X.12349.

[9] Chao, A., Kubota, Y., Zelený, D., Chiu, C.-H., Li, C.-F., Kusumoto, B., Yasuhara, M., et al. (2020). Quantifying sample completeness and comparing diversities among assemblages. *Ecological Research*, 35, 292–314. doi:10.1111/1440-1703.12102.

[10] Chisholm, R. A. (2026). The usefulness of evenness. *Ecology Letters*, 29, e70446. doi:10.1111/ele.70446.

[11] Chiu, C.-H., Wang, Y.-T., Walther, B. A., and Chao, A. (2014). An improved nonparametric lower bound of species richness via a modified Good–Turing frequency formula. *Biometrics*, 70, 671–682. doi:10.1111/biom.12200.

[12] Hill, M. O. (1973). Diversity and evenness: a unifying notation and its consequences. *Ecology*, 54, 427–432.

[13] Jost, L. (2010). The relation between evenness and diversity. *Diversity*, 2, 207–232.

[14] Pielou, E. C. (1966). The measurement of diversity in different types of biological collections. *Journal of Theoretical Biology*, 13, 131–144.

[15] Sharma, B. D., and Mittal, D. P. (1975). New non-additive measures of entropy for discrete probability distributions. *Journal of Mathematical Sciences (Delhi)*, 10, 28–40.

[16] Smith, B., and Wilson, J. B. (1996). A consumer's guide to evenness indices. *Oikos*, 76, 70–82. doi:10.2307/3545749.