

How Many Correlated Signals Actually Diversify? A Controlled Validation of Effective-Breadth Approximations for Trading-Signal Portfolios

Eugen Soloviov*

Abstract

Running one strategy across N instruments is often assumed to give N -fold diversification, but correlated signals deliver far less. A popular shortcut borrows the equicorrelation *effective sample size* $N_{\text{eff}} = N/(1+(N-1)\bar{\rho})$ from statistics and uses it both to discount diversification and to size the capacity of a slot-limited execution orchestrator. We give this shortcut a controlled test. In a simulation with known population correlation we generate N correlated *binary* trading signals from a latent Gaussian model, attach trade returns whose loading β on the common factor is an explicit design parameter swept over $\{0, 0.3, 0.6, 0.9\}$, and compare N_{eff} against two ground truths: the realized variance reduction of the equal-weight portfolio and a correlated-Bernoulli model of execution capacity. Across 4,000 simulated portfolios we find: (i) for the binary signal streams themselves N_{eff} is essentially exact (0.02% mean error); (ii) for the *PnL* streams its accuracy—and even the sign of its bias—is set by the return model: fed the binary signal correlation, the bias swings from -56% at $\beta = 0$ to $+91\%$ at $\beta = 0.9$ ($+22\%$ at the central $\beta = 0.6$). What is robust across β is the *feeding gap*: dichotomization attenuates correlation by roughly half (a tetrachoric effect we predict to 0.018 MAE), so the binary-fed estimate always claims about 56% more breadth than the latent-fed one; (iii) the capacity heuristic $1 - (1 - p)^{N_{\text{eff}}}$ misses a structural fact—the expected number of simultaneously active signals is Np by linearity *regardless* of correlation—and systematically undershoots $P(\geq 1 \text{ active})$. Fed the latent correlation its error (MAE 0.21) is worse than ignoring correlation altogether (0.14); fed the binary correlation it improves to 0.11 and beats the naive model in 68% of portfolios. Both feedings bias the implied “optimal number of instruments” low (by 4–6 and 2–4 at our homogeneous design point), though the flat score curve makes the realized throughput cost modest ($\leq 7\%$). We conclude that N_{eff} is a usable breadth discount only when the fed correlation matches the stream one cares about, and that slot capacity should be sized from Np and the distribution of the active count, not from N_{eff} .

1 Introduction

A trader who runs the same strategy on N instruments would like to believe the result is N roughly-independent bets. It is not: the signals fire together, the positions move together, and the realized diversification is a fraction of the nominal count [5, 10]. The practical question is how large that fraction is, and whether it can be captured by a formula simple enough to use on the desk.

*Independent Researcher. ORCID: 0009-0006-3148-111X. Correspondence: suenot@gmail.com. Code to reproduce every number and figure: <https://github.com/suenot/signal-breadth>.

One such formula circulates widely. Borrowed from the design-effect of survey sampling [7], the equicorrelation *effective sample size*

$$N_{\text{eff}} = \frac{N}{1 + (N - 1)\bar{\rho}} \quad (1)$$

discounts N by an average pairwise correlation $\bar{\rho}$. It is attractive because it needs only N and one number, sidestepping the noisy full-covariance estimation that more principled measures require [8, 9]. In practitioner writing it is used for two distinct purposes: to estimate how much diversification a basket of correlated strategies actually delivers, and—plugged into an independent-arrivals formula $1 - (1 - p)^{N_{\text{eff}}}$ —to size how busy a capacity- or slot-limited execution system will be.¹ Neither use has, to our knowledge, been validated against ground truth.

We provide that validation in a controlled simulation. Because we generate the data, we know the population correlation, the oracle diversification, and the true distribution of simultaneously-active signals, and can measure exactly where (1) succeeds and where it fails. Our findings are deliberately mixed: the formula is far better than its reputation on one target, far worse on another, and the difference is governed by two channels the shortcut ignores—which correlation it is fed, and how strongly returns co-move beyond co-activation.

Contributions.

1. A reproducible framework that generates N correlated *binary* trading signals and their PnL from a latent Gaussian model with controllable homogeneous or heterogeneous correlation *and an explicit return-factor loading* β , so realized breadth and execution capacity are measurable ground truth and the return channel is a swept design parameter rather than a hidden constant (Section 4).
2. A measurement of N_{eff} 's accuracy as a diversification discount. For the signal streams themselves it is essentially exact (0.02% error). For the PnL streams, accuracy and the *sign* of the bias depend on β ; the β -robust feature is the feeding gap driven by tetrachoric attenuation, which makes the binary-fed estimate claim $\sim 56\%$ more breadth than the latent-fed one (Sections 5.1–5.3).
3. A clarification that Meucci's Effective Number of Bets, though principled, measures factor breadth rather than equal-weight variance reduction and is not a drop-in ground truth for this problem (Section 5.4).
4. A precise refutation of the capacity use: the mean number of active signals is Np by linearity for *any* correlation (the $N_{\text{eff}} p$ load estimate is structurally wrong), and the $P(\geq 1)$ heuristic systematically undershoots, with an error whose size—and rank relative to a naive independence model—depends on which correlation is fed. The implied optimal instrument count is biased low by 2–6, at a modest ($\leq 7\%$) throughput cost (Sections 5.5–5.6).

2 Related work

The question of how many genuinely distinct bets a portfolio contains is as old as modern portfolio theory. Markowitz [10] formalized diversification through the covariance structure: portfolio

¹The specific practitioner formulation we test, including the $1 - (1 - p)^{N_{\text{eff}}}$ capacity step and the $\mathbb{E}[\text{active}] = N_{\text{eff}} p$ load estimate, originates in the author's own earlier practitioner writing (a market-making blog post). This paper is therefore a controlled self-audit of advice the author has himself circulated, not an attack on a third party.

variance depends not on the number of holdings but on their co-movement, so the naive count of N positions overstates diversification whenever holdings are correlated. We study the analogous overstatement for N *trading signals* and ask whether a closed-form correction recovers the effective count.

A literature seeks scalar summaries of diversification. Meucci [11] proposed the Effective Number of Bets (ENB): rotate the portfolio onto its uncorrelated principal portfolios, normalize their risk contributions into a diversification distribution, and exponentiate its Shannon entropy. The ENB equals N when risk is spread evenly over N uncorrelated sources and collapses toward one as risk concentrates. In this paper we use the spectral, portfolio-agnostic variant of Meucci’s measure—the exponential entropy of the correlation-matrix eigenvalue shares—which depends only on the correlation structure, not on a particular weight vector. It is correlation-aware and principled, but requires the full covariance matrix and an eigendecomposition—and, as the random-matrix critique [8] and shrinkage remedy [9] stress, that matrix is itself noisily estimated.

The active-management literature attacks the same problem through *breadth*. Grinold [5] introduced the Fundamental Law, $IR \approx IC \cdot \sqrt{BR}$, with breadth BR the number of *independent* bets per period; Grinold and Kahn [6] develop it in detail. Independence is the load-bearing assumption: when signals are correlated the raw count overstates breadth. Clarke et al. [1] generalized the law to constrained portfolios via the transfer coefficient, and Clarke et al. [2] extended it to a full covariance matrix, making explicit that breadth shrinks and becomes hard to define once signals are correlated. These works establish qualitatively that correlation erodes effective breadth, but supply no validated closed-form effective count for an arbitrary number of binary signals.

A candidate formula exists in classical statistics: under equicorrelation the variance of the sample mean of N unit-variance variables is inflated by $1 + (N - 1)\bar{\rho}$, giving the effective sample size (1), the design-effect of survey sampling [7]. Two features of trading signals complicate its transplant. First, signals are typically *binary*, and the relevant correlation is between Bernoulli indicators, not their latent drivers; dichotomizing a bivariate normal attenuates the observed correlation in a non-linear, base-rate-dependent way—the classical tetrachoric problem [12]. Second, dependence is not fully captured by linear correlation: stylized facts [3] and copula critiques [4] warn that a single $\bar{\rho}$ can mislead. A validation must therefore be explicit about which correlation it uses and under what structure it is asked to hold. We close this gap by simulation, taking the realized variance reduction and a correlated-Bernoulli capacity model as ground truth.

3 Effective-breadth measures

We compare three quantities. The **equicorrelation** N_{eff} of (1), the object under test, defined for any average correlation $\bar{\rho}$ in its positive-semidefinite range $\bar{\rho} \geq -1/(N - 1)$ (a constraint the shortcut routinely ignores; we enforce it, though it never binds for our $\rho \geq 0$ designs). Meucci’s **Effective Number of Bets**,

$$\text{ENB} = \exp\left(-\sum_{k=1}^N p_k \ln p_k\right), \quad p_k = \frac{\lambda_k}{\sum_j \lambda_j}, \quad (2)$$

with λ_k the eigenvalues of the correlation matrix—a principled measure of the number of uncorrelated risk *factors*. And the operational ground truth, the **realized effective** N , defined as the variance-reduction ratio of the equal-weight portfolio of the N signal-PnL streams,

$$N^* = \frac{\overline{\text{Var}}(\text{single stream})}{\text{Var}(\text{equal-weight portfolio})}, \quad (3)$$

which for *equal-variance* equicorrelated streams equals $N/(1 + (N - 1)\bar{\rho}_{\text{PnL}})$ exactly, and is what a trader actually experiences as diversification.

Latent vs. binary correlation. A signal is active when a latent driver $z_{i,t} = \sqrt{\rho} F_t + \sqrt{1 - \rho} \varepsilon_{i,t}$ exceeds a threshold set to give activation probability p . The correlation of the resulting binary indicators is *not* ρ ; thresholding attenuates it [12]. Whether one feeds N_{eff} the binary signal correlation or the latent correlation therefore changes the answer, and the shortcut is silent on which to use. We evaluate both feedings throughout.

4 Simulation framework

Each experiment draws a population of N signals over T periods (Figure 1). Latent drivers follow a one-factor Gaussian model; in the *homogeneous* case all loadings are equal (true equicorrelation ρ), while in the *heterogeneous* case loadings are dispersed and an optional block structure (2–3 sector factors) is added, so pairwise correlations are unequal and (1)’s single- $\bar{\rho}$ assumption is stressed. Loading rows are normalized so every latent driver has exactly unit variance, which guarantees the realized activation probability matches the configured p . Signal i is active when its latent driver exceeds $\Phi^{-1}(1 - p)$.

Return model. Each active step of signal i earns

$$r_{i,t} = \mu + \sigma(\beta F_t + \sqrt{1 - \beta^2} \eta_{i,t}), \quad (4)$$

with edge $\mu = 0.05$, volatility $\sigma = 1$, F_t the *same* common factor that drives activations, and $\eta_{i,t}$ idiosyncratic noise; flat steps earn zero. The loading β controls how much of the co-activation structure passes through to PnL correlation: at $\beta = 0$ the PnL streams are correlated only through co-activation of the common edge term, at $\beta = 0.9$ they co-move strongly. Because the validation target—realized PnL variance reduction—depends directly on β , we treat it as an explicit design parameter and sweep it, rather than fixing it implicitly.

Sampled design. The main batch draws, independently per experiment: $N \in \{5, 8, 10, 15, 20, 30\}$, $p \in \{0.05, 0.10, 0.15, 0.25, 0.45\}$, $\rho \sim \text{Uniform}[0, 0.7]$ (clipped to the PSD-valid range, which never binds for $\rho \geq 0$), $\beta \in \{0, 0.3, 0.6, 0.9\}$, and homogeneous vs. heterogeneous structure with equal probability; heterogeneous configs draw a loading dispersion in $[0.1, 0.45]$ and, with probability 2/3, 2–3 blocks carrying up to 25% of variance. We run 4,000 portfolios of $T = 30,000$ periods each (1,995 homogeneous, 2,005 heterogeneous; about 1,000 per β level). The ρ -sweep of Section 5.5 holds $N = 10$, $p = 0.1$ fixed with $T = 120,000$ and five repetitions per ρ ; the optimal- N exercise of Section 5.6 uses $T = 60,000$, three repetitions per candidate N , and five independent seeds per setting. All runs are deterministic given the released seeds. For execution we add an orchestrator with K slots: at each step at most K of the active signals can be filled.

Ground truth is read off the simulation: the realized effective N above; the empirical distribution of the number of simultaneously active signals; and the slot utilization $\mathbb{E}[\min(\text{active}, K)]/K$. We then ask how well the N_{eff} -based predictions match.

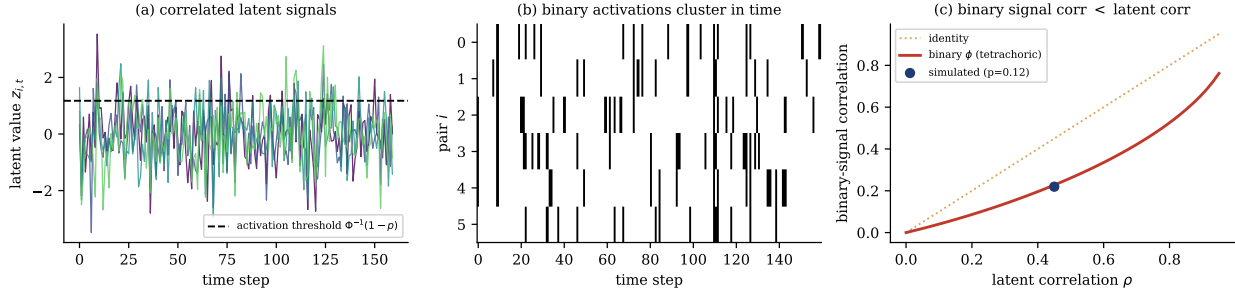


Figure 1: The generative model. (a) Correlated latent Gaussian drivers and the activation threshold $\Phi^{-1}(1 - p)$. (b) The resulting binary activations cluster in time. (c) Dichotomization attenuates correlation: the binary (ϕ) correlation lies well below the latent correlation that produced it; the tetrachoric curve predicts the attenuation.

Table 1: Accuracy of effective-breadth estimators against the realized variance-reduction effective N of the PnL streams (4,000 portfolios, marginal over the β grid). MAPE = mean absolute percentage error; bias is signed.

Estimator	Overall		Homogeneous		Heterogeneous	
	MAPE	bias	MAPE	bias	MAPE	bias
N_{eff} (binary corr.)	51.7%	+4.1%	51.8%	+2.1%	51.5%	+6.1%
N_{eff} (latent corr.)	49.3%	-28.4%	49.5%	-31.4%	49.0%	-25.5%
Meucci ENB (latent)	103.9%	+85.1%	108.6%	+89.4%	99.2%	+80.8%

5 Results

5.1 As a PnL diversification discount, N_{eff} 's bias is set by the return channel

Figure 2 and Tables 1–2 report how well N_{eff} predicts the realized PnL effective N . Marginally over the β grid the binary-fed formula shows a mean absolute error of 51.7% (bias +4.1%) and the latent-fed one 49.3% (bias -28.4%)—but the marginal numbers conceal the real structure, which is the β -dependence in Table 2. At $\beta = 0$, where returns share no factor beyond co-activation, realized PnL breadth stays close to N and *both* feedings under-count badly (-56% binary-fed, -69% latent-fed). At $\beta = 0.9$ both over-count (+91% and +30%). The sign of the bias flips between: at the central $\beta = 0.6$ the binary feeding over-counts (+21.7%, MAPE 25.8%) while the latent feeding under-counts (-16.1%, MAPE 32.6%). A signal-level correlation simply does not contain the information (β) that determines PnL co-movement, so no feeding of (1) can be unbiased across return regimes.

Two features *are* robust across β . First, the ordering: the binary-fed estimate always claims more breadth than the latent-fed one, by a construction-driven $\sim 56\%$ (next subsection), with the bias spread between the two feedings widening from 12.6pp at $\beta = 0$ to 61.4pp at $\beta = 0.9$ as the realized breadth it is measured against shrinks. Second, structure barely matters: accuracy degrades little from homogeneous (51.8%) to heterogeneous (51.5%) correlation—the binary attenuation and the return channel, not correlation heterogeneity, dominate the error.

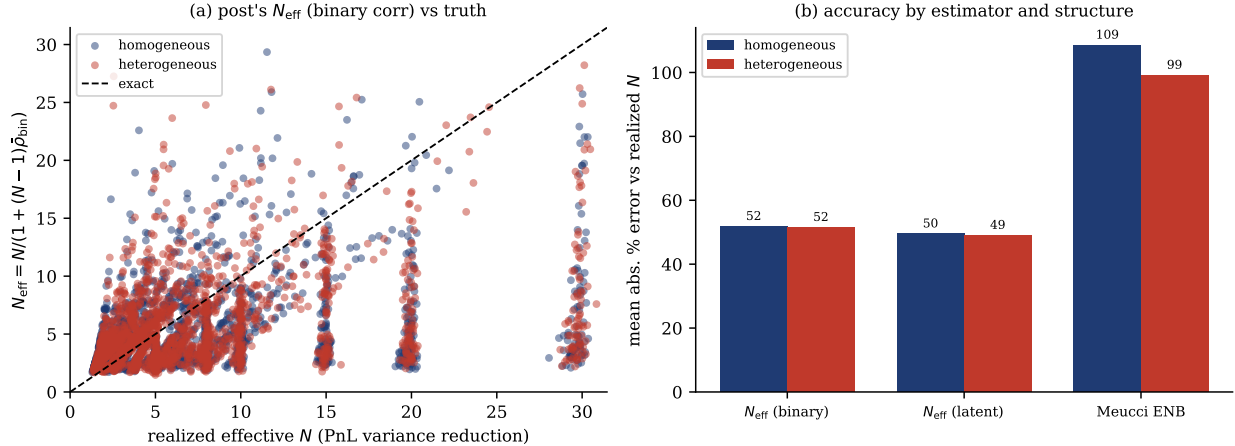


Figure 2: (a) The shortcut’s N_{eff} (fed binary correlation) versus the realized PnL effective N , all 4,000 portfolios shown (axes span the full data range). The vertical bands near $N^* \approx N$ are low- β portfolios whose PnL streams are nearly independent. (b) Mean absolute percentage error by estimator and structure, marginal over the β grid. Meucci’s ENB measures a different quantity and diverges far more (Section 5.4).

Table 2: The return-factor loading β sets the size *and sign* of N_{eff} ’s bias against realized PnL breadth ($\sim 1,000$ portfolios per row). The feeding gap (binary minus latent bias) is the tetrachoric-driven, β -robust ordering; its magnitude in percentage points grows with β because the realized breadth in the denominator shrinks.

β	N_{eff} (binary corr.)		N_{eff} (latent corr.)		
	MAPE	bias	MAPE	bias	spread (pp)
0.0	56.4%	−56.4%	69.0%	−69.0%	12.6
0.3	35.7%	−33.7%	55.3%	−54.3%	20.6
0.6	25.8%	+21.7%	32.6%	−16.1%	37.7
0.9	91.4%	+91.4%	39.3%	+30.0%	61.4

5.2 The binary/latent distinction is large and predictable

The reason *which* correlation one feeds matters is that dichotomizing attenuates it. Across our portfolios the mean latent correlation of 0.324 becomes a mean binary signal correlation of 0.176—a median per-portfolio attenuation ratio of 0.52, with the binary correlation below the latent one in 98% of cases. The classical tetrachoric relation predicts the binary correlation from the latent one and the activation rate to a mean absolute error of just 0.018. The practical consequence is direct: feeding N_{eff} the (smaller) binary correlation instead of the latent one inflates the estimated effective count by +56% on average (5.19 vs. 3.43). This inflation depends only on the activation geometry, not on β —which is exactly why the binary-fed estimate sits above the latent-fed one at every β in Table 2. A user of (1) who measures correlation on the signals themselves—the natural thing to do—always claims more diversification than a latent-correlation user, and overstates the truth whenever return co-movement is appreciable.

5.3 The formula is near-exact for the signal streams; the PnL error is the return channel

Two further measurements locate the error. First, judged against the variance reduction of the *binary signal streams themselves*, the binary-fed N_{eff} is essentially exact: 0.022% MAPE across all 4,000 portfolios, heterogeneous and block structures included. This is close to an algebraic identity—for equal-variance streams the equal-weight variance-reduction ratio depends on the correlation matrix only through its mean—so it should be read as a consistency check, but a useful one: it rules out any contribution of correlation *heterogeneity* per se to the PnL-stream error. The entire $\sim 50\%$ PnL error of Table 1 comes from the mismatch between signal-level and PnL-level correlation, i.e. from tetrachoric attenuation and the return channel β , not from the equicorrelation algebra.

Second, the error is concentrated in the weak-correlation regime. At the central $\beta = 0.6$, the binary-fed MAPE falls from 62% in the lowest- ρ bin ($\rho \leq 0.1$) to 9% for $\rho \in (0.45, 0.7]$: when correlation is strong it dominates both the signal and the PnL structure, and the input mismatch matters less. (The direction of this ρ -dependence is itself β -conditional: at $\beta = 0$, where realized PnL breadth stays near N , the formula’s aggressive discounts make the error *grow* with ρ instead, from 23% to 76%.) The uncomfortable reading for practice: in the regime where a correlation correction seems most affordable—weak-to-moderate measured correlation—the shortcut is least reliable, and where it is most accurate the correction is largest.

5.4 Meucci’s ENB answers a different question

It is tempting to treat Meucci’s ENB as the “true” effective number and judge N_{eff} against it. Our results caution against this. The ENB diverges from the realized equal-weight variance reduction by 104% (computed on latent correlations) to 120% (on return correlations), far more than N_{eff} does (Table 1). This is not a failure of the ENB: it faithfully measures the number of uncorrelated risk *factors* (the entropy of the eigenvalue spectrum), which is a different quantity from the variance reduction obtained by a specific equal-weight combination. The lesson is that “effective number” is not one quantity, and a capacity or diversification claim must name the operational target. For the equal-weight signal basket that motivates the shortcut, the variance-reduction N^* is the relevant ground truth, and N_{eff} —for all its roughness on PnL streams—is far closer to it than the spectral ENB.

5.5 As a capacity model: the mean load is structurally wrong, and the $P(\geq 1)$ error depends on the feeding

The second use of N_{eff} is to size execution capacity via $1 - (1 - p)^{N_{\text{eff}}}$ and $\mathbb{E}[\text{active}] = N_{\text{eff}} p$. The mean-load identity is wrong by construction: by linearity of expectation the mean number of simultaneously active signals is Np for *any* correlation structure—correlation changes the *variance* of the active count, not its mean. Empirically, Np predicts the simulated mean active count to a MAE of 0.011 (sampling noise), whereas $N_{\text{eff}} p$ is off by 2.13 latent-fed (recovering $0.31\times$ the true load) and 1.87 binary-fed ($0.44\times$; Table 3).

The $P(\geq 1 \text{ active})$ heuristic is wrong more subtly: clustering does reduce the union probability, but less than the heuristic claims, so it systematically undershoots—and by how much depends on which correlation is fed. Latent-fed, the MAE is 0.209 (bias -0.209), worse than even the naive independent- N model (0.137); it beats the naive model in only 20% of portfolios. Binary-fed—the input a practitioner measuring the signals would actually use—the attenuated correlation yields a larger N_{eff} and a prediction much closer to the truth: MAE 0.105, beating the naive model in 68%

Table 3: Execution-capacity predictions vs. simulation under both feedings of N_{eff} . The reference column gives the true Np for the mean-load rows and the naive independent model $1 - (1 - p)^N$ for the $P(\geq 1)$ row.

Quantity	N_{eff} latent-fed	N_{eff} binary-fed	reference
Mean active count, MAE	2.13	1.87	0.011 (truth Np)
Mean active count, ratio to truth	$0.31\times$	$0.44\times$	$1.00\times$ (truth Np)
$P(\geq 1)$, $K=1$, MAE	0.209	0.105	0.137 (naive)
Slot utilization $K=3$ (sim 0.491)	0.226	0.306	—

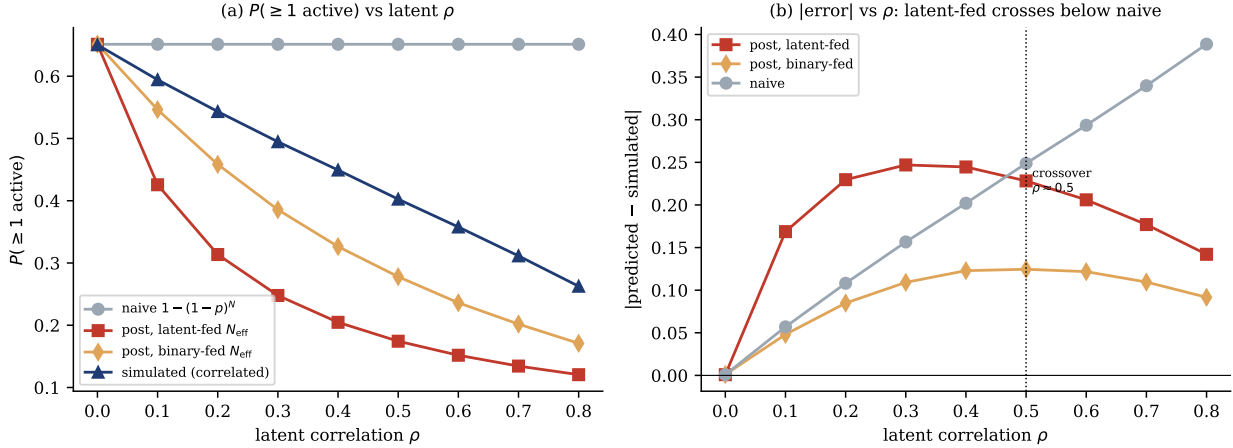


Figure 3: Execution-capacity error of $1 - (1 - p)^{N_{\text{eff}}}$ ($N = 10$, $p = 0.1$, five repetitions per ρ). (a) Both feedings undershoot the simulated $P(\geq 1 \text{ active})$; the binary-fed prediction lies between the latent-fed one and the naive model. (b) Absolute errors: the latent-fed error peaks near $\rho \approx 0.3$ and crosses *below* the naive model’s error at $\rho \approx 0.5$ —at high correlation the heuristic beats ignoring correlation—while the binary-fed error is below the naive error for all $\rho \geq 0.1$. The error does not grow monotonically with ρ ; what grows fastest is the naive model’s error.

of portfolios. The undershoot is structural in both cases, but the common practice of quoting the latent-fed numbers alone overstates the indictment.

The ρ -dependence must also be stated precisely. The *absolute error* of both feedings rises with $\bar{\rho}$ in the main batch (Spearman rank correlation 0.49 latent-fed and 0.65 binary-fed, both $p < 10^{-200}$), but the naive model’s error rises faster. In the controlled sweep ($N = 10$, $p = 0.1$; Figure 3) the latent-fed error peaks near $\rho \approx 0.3$ and *crosses below the naive error* at $\rho \approx 0.5$: at high correlation the heuristic beats ignoring correlation, despite its bias. The binary-fed error sits below the naive error for all $\rho \geq 0.1$. By $\bar{\rho}$ -tercile of the main batch, the latent-fed / binary-fed / naive MAEs are 0.13/0.04/0.04 (low), 0.25/0.12/0.13 (mid), and 0.25/0.15/0.24 (high). With three slots, true utilization is 0.491; the heuristic predicts 0.226 latent-fed and 0.306 binary-fed. The heuristic conflates a real effect (correlation makes the active count *lumpier*, so a fixed slot budget is sometimes overwhelmed and often idle) with a fall in the *mean* load, which does not occur.

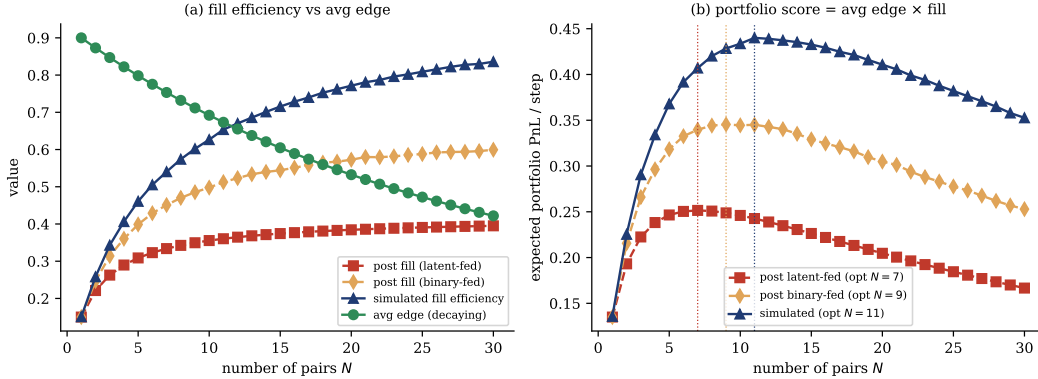


Figure 4: Throughput vs. number of instruments under a slot budget and decaying edge (homogeneous design, representative seed). The latent-fed heuristic’s score peaks at $N = 7$ and the binary-fed at $N = 9$, vs. the simulated optimum $N = 11$ in this seed (median 12, range 11–13 across five seeds). The simulated score curve is flat near its maximum, so the throughput cost of the early stop is $\sim 7\%$ (latent-fed) and $\sim 2\%$ (binary-fed).

5.6 The implied “optimal number of pairs” is biased low—at a modest cost

Putting the pieces together, a common application chooses the number of instruments N that maximizes throughput under a slot budget and a decaying per-instrument edge. Because the N_{eff} capacity model undershoots fill rates, it stops adding instruments too early. Under the homogeneous design point (one slot, $p = 0.15$, $\rho = 0.3$, decaying edge), across five independent seeds the simulated optimum is $N_{\text{sim}}^* = 11\text{--}13$ (median 12), while the latent-fed heuristic recommends $N = 7$ in every seed (gap -4 to -6) and the binary-fed variant $N = 9\text{--}10$ (gap -2 to -4). Under a genuinely heterogeneous design (dispersed loadings, three blocks), the simulated optimum is $10\text{--}15$ (median 13) against latent-fed recommendations of $7\text{--}13$ (gap -7 to 0) and binary-fed $8\text{--}15$ (gap -5 to 0).

Two qualifications temper the headline. First, the simulated optimum itself moves by ± 2 across seeds, so single-seed gaps (including the 12-vs-13 homogeneous-vs-heterogeneous contrast in an earlier version of this experiment) are partly seed noise; we therefore report ranges. Second, the score curve is flat near its maximum (Figure 4b): trading at the latent-fed recommendation forfeits on average only 6.9% (homogeneous) and 5.1% (heterogeneous) of attainable throughput, and the binary-fed recommendation only 1.7% and 1.1%. The heuristic stops early, but the cliff it stops before is shallow. A desk following the formula under-deploys, mildly.

6 Discussion

The shortcut $N_{\text{eff}} = N / (1 + (N - 1)\bar{\rho})$ should be split into its two uses, and each verdict must name the correlation being fed.

As a *diversification discount* for the signal streams themselves it is essentially exact, and as a PnL discount it is rough at best: its bias is set by the return-factor loading β , which no signal-level correlation contains. Neither feeding is reliably unbiased—the binary-fed version’s near-zero marginal bias ($+4\%$) is an artifact of opposite-sign errors cancelling across return regimes (Table 2). The defensible default is the *latent* (or return) correlation, read as a conservative discount: in our grid it under-counts breadth everywhere except the extreme $\beta = 0.9$ corner, and under-counting is the safe direction for risk. Feeding the binary signal correlation claims $\sim 56\%$ more breadth

by construction and overstates true diversification precisely when return co-movement is strong—the dangerous direction. Either way the result should be judged against equal-weight variance reduction, not the spectral ENB, which measures factor breadth.

As a *capacity model* the verdict is harsher but should be precise rather than blanket. The mean-load identity $\mathbb{E}[\text{active}] = N_{\text{eff}} p$ is simply wrong: the truth is Np irrespective of correlation, and the relevant effect of correlation is on the *dispersion* of the load, which a correlated-Bernoulli (or directly simulated) model captures and $1 - (1 - p)^{N_{\text{eff}}}$ does not. The $P(\geq 1)$ heuristic undershoots systematically at low-to-mid $\bar{\rho}$, and its quality depends on the feeding: latent-fed it is worse than assuming independence; binary-fed it is better than independence in two out of three portfolios, and its optimal- N recommendation costs only $\sim 1\text{--}2\%$ of throughput. If the heuristic is used at all for fill estimates, it should be fed the binary correlation; load and slot budgets should be sized from Np and the distribution of the active count.

7 Limitations

Our dependence is Gaussian-copula with one factor plus optional blocks; real signals exhibit tail co-movement and regime-dependent correlation [3, 4] that a single $\bar{\rho}$ cannot represent and that would further stress N_{eff} . Our return model (4) ties PnL co-movement to a single constant β per portfolio; real return loadings vary across instruments and time, and our β grid (0–0.9), while spanning weak to strong co-movement, is itself a modeling choice. We study equal-weight baskets and a simple slot orchestrator; capacity-aware weighting, queueing, and priority among signals are out of scope. Activation is a fixed-threshold binary; graded position sizing would change the tetrachoric attenuation. Finally we model no transaction costs, and our “edge decay” in the optimal- N exercise is a stylized assumption, not an empirical estimate. None of these affects the central structural result— $\mathbb{E}[\text{active}] = Np$ —which is exact.

8 Conclusion

We tested a popular effective-number shortcut for portfolios of correlated trading signals against simulated ground truth. As a diversification discount it is essentially exact for the binary signal streams themselves, but for PnL streams its accuracy—including the sign of its bias—depends on how strongly returns co-move beyond co-activation, a quantity (β) that no signal-level correlation reveals; the robust, predictable effect is tetrachoric attenuation, which makes any binary-fed estimate claim $\sim 56\%$ more breadth than its latent-fed counterpart. As a model of execution capacity its mean-load arithmetic is structurally wrong—the mean load is the correlation-free Np —and its $P(\geq 1)$ estimate undershoots, badly when fed the latent correlation (worse than assuming independence) and tolerably when fed the binary correlation, biasing the implied optimal instrument count low by 2–6 at a 1–7% throughput cost. Use N_{eff} as a conservative breadth discount fed the latent or return correlation; size slot budgets from Np and the distribution of the active count, not from N_{eff} .

Reproducibility. All experiments are deterministic given the released seeds; the script `run_all.py` regenerates every number and figure, and `scripts/check_paper_numbers.py` verifies that every number quoted in this paper matches the generated results. The generative model, the breadth and capacity measures, and the analysis are provided as an open-source package.

References

- [1] Roger Clarke, Harindra de Silva, and Steven Thorley. Portfolio constraints and the fundamental law of active management. *Financial Analysts Journal*, 58(5):48–66, 2002. doi: 10.2469/faj.v58.n5.2468.
- [2] Roger Clarke, Harindra de Silva, and Steven Thorley. The fundamental law of active portfolio management. *Journal of Investment Management*, 4(3), 2006.
- [3] Rama Cont. Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236, 2001. doi: 10.1080/713665670.
- [4] Paul Embrechts, Alexander J. McNeil, and Daniel Straumann. Correlation and dependence in risk management: Properties and pitfalls. In M. A. H. Dempster, editor, *Risk Management: Value at Risk and Beyond*, pages 176–223. Cambridge University Press, Cambridge, 2002. doi: 10.1017/CBO9780511615337.008.
- [5] Richard C. Grinold. The fundamental law of active management. *The Journal of Portfolio Management*, 15(3):30–37, 1989. doi: 10.3905/jpm.1989.409211.
- [6] Richard C. Grinold and Ronald N. Kahn. *Active Portfolio Management: A Quantitative Approach for Producing Superior Returns and Controlling Risk*. McGraw-Hill, New York, 2 edition, 2000. ISBN 978-0-07-024882-3.
- [7] Leslie Kish. *Survey Sampling*. John Wiley & Sons, New York, 1965.
- [8] Laurent Laloux, Pierre Cizeau, Jean-Philippe Bouchaud, and Marc Potters. Noise dressing of financial correlation matrices. *Physical Review Letters*, 83(7):1467–1470, 1999. doi: 10.1103/PhysRevLett.83.1467.
- [9] Olivier Ledoit and Michael Wolf. Honey, I shrunk the sample covariance matrix. *The Journal of Portfolio Management*, 30(4):110–119, 2004. doi: 10.3905/jpm.2004.110.
- [10] Harry Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952. doi: 10.1111/j.1540-6261.1952.tb01525.x.
- [11] Attilio Meucci. Managing diversification. *Risk*, 22(5):74–79, 2009. Also available as SSRN Working Paper 1358533.
- [12] Karl Pearson. Mathematical contributions to the theory of evolution. VII. on the correlation of characters not quantitatively measurable. *Philosophical Transactions of the Royal Society of London, Series A*, 195:1–47, 1900. doi: 10.1098/rsta.1900.0022.