

Two Sides of Schur Damping

High-Dimensional Pseudo-Likelihoods and Portfolio Allocation

Peter Cotton

Note — microprediction/precise

Abstract

Two communities that rarely cite each other—spatial statisticians fitting high-dimensional weather fields, and quantitative investors building portfolios—have independently arrived at the same mathematical object: a Schur complement, damped by one interpretable parameter. In spatial modeling the Schur complement is the *conditional covariance* that makes a Gaussian (Vecchia) pseudo-likelihood estimable at scale, and recent work regularizes it by shrinking toward a base model. In allocation it is the *residual risk* of a bet net of its hedge, and the same parameter interpolates hierarchical risk parity and the minimum-variance portfolio. We show these are one operation—reliability shrinkage of a conditional Gaussian—so that the damping a weather model needs to remain estimable when stations outnumber observations is, term for term, the damping a portfolio needs to remain stable when assets outnumber returns. The optimal amount is a closed-form reliability, a James–Stein shrinkage that is simultaneously a Ledoit–Wolf intensity. The shrinkage machinery is classical, but the *identity* appears to be new: to our knowledge neither literature has noted that the conditional shrinkage a spatial model fits and the diversification–variance tilt a portfolio chooses are one and the same quantity. We make the correspondence precise, note that the two literatures have each supplied what the other lacks, and report a small experiment on the one genuinely open choice—how to set the damping—suggesting the spatial community’s fitted intensity is, if anything, the better recipe.

1 An unlikely correspondence

A meteorologist estimating a temperature field over tens of thousands of stations and a portfolio manager allocating across thousands of assets face, superficially, unrelated problems. Yet both are defeated by the same thing: a covariance matrix whose dimension p rivals or exceeds the number of observations n , so that its inverse—needed to condition one variable on the others, or to weight one asset against the others—is unstable or undefined. And both fields have converged on the same fix, a Schur complement damped by a single parameter, without noticing they agree.

On the **weather** side, scalable Gaussian-process inference uses the Vecchia factorization [Vecchia, 1988, Katzfuss and Guinness, 2021], which conditions each location only on a few neighbours; the conditional covariances are Schur complements, and Chakraborty and Katzfuss [2025] (ShrinkTM), building on the Bayesian transport maps of Katzfuss and Schäfer [2024], improve small-sample behaviour by *shrinking* those conditionals toward a parametric base, learning the strength by empirical Bayes.

On the **portfolio** side, Schur-complementary allocation—introduced in 2022 [Cotton, 2022] and developed in Cotton [2024]—damps the off-diagonal coupling of the covariance through its Schur complements by a parameter γ , recovering hierarchical risk parity [López de Prado, 2016] at $\gamma = 0$ and the minimum-variance portfolio at $\gamma = 1$. The dial thus predates the spatial shrinkage work below by some two years, the two having developed independently.

This note shows the two are the same Schur damping, with the same closed-form optimal γ , and that each community has supplied a piece the other was missing.

2 The shared object: one Schur complement, two readings

Let R be a correlation matrix and partition the variables into a block k and a conditioning set c (its spatial neighbours, or the other assets). The Schur complement

$$S_k = R_{kk} - R_{kc}R_{cc}^{-1}R_{ck} \quad (1)$$

carries the same algebra into two meanings.

Weather (a pseudo-likelihood term). S_k is the covariance of block k *conditional* on c , and $b_k = R_{cc}^{-1}R_{ck}$ is the regression of k on its neighbours. The Gaussian density factorizes as $\prod_k \mathcal{N}(y_k; b_k^\top y_c, S_k)$; truncating c to m nearest neighbours is the Vecchia pseudo-likelihood, the workhorse for fitting high-dimensional spatial fields. (This S_k, b_k are exactly the τ^2, ξ of Chakraborty and Katzfuss, 2025, Eq. 8.)

Portfolio (a residual risk). S_k is the covariance of block k *net of the best hedge* formed from c —the risk that remains after the rest of the book offsets it—and b_k is that hedge. Splitting capital by inverse residual risk is precisely how Schur-complementary allocation [Cotton, 2024] interpolates between treating blocks independently (risk parity) and fully exploiting their cross-hedging (minimum variance).

Same S_k , same b_k : a conditional variance for the statistician, a hedged residual risk for the investor.

A two-variable example. Take two standardized variables with correlation ρ . The Schur complement (1) is the scalar $S = 1 - \rho^2$. To the statistician it is the variance of variable 2 conditional on variable 1, $\text{Var}(y_2 | y_1) = 1 - \rho^2$ —the term the pseudo-likelihood scores. To the investor it is the variance of variable 2 *net of its hedge* by variable 1: with optimal hedge ratio $b = \rho$, the residual $y_2 - b y_1$ has variance $1 - \rho^2$, the risk that remains after the hedge. Identical number, two stories. Damping (2) gives $S(\gamma) = (1 - \gamma) \cdot 1 + \gamma(1 - \rho^2) = 1 - \gamma \rho^2$: at $\gamma = 1$ the full hedge / full conditioning, at $\gamma = 0$ none (the marginal variance, an unhedged position), and at $\gamma = \gamma^*$ a residual risk that trusts the estimated coupling only as far as it is reliable. The pseudo-likelihood and the portfolio read the same $1 - \gamma \rho^2$ off the same complement.

3 The shared cure: damping by reliability

Both readings break for the same reason— R_{cc}^{-1} is estimated from limited data, so b_k and S_k overfit—and both apply the same cure, a convex damping by $\gamma \in [0, 1]$:

$$b_k(\gamma) = \gamma b_k, \quad S_k(\gamma) = (1 - \gamma) R_{kk} + \gamma S_k. \quad (2)$$

At $\gamma = 1$ this is full conditioning—the exact Gaussian likelihood, the minimum-variance portfolio. At $\gamma = 0$ it ignores the coupling—the block-diagonal composite likelihood, hierarchical risk parity. Intermediate γ trusts the estimated cross-coupling only partway. The weather and portfolio extremes are the *same* two endpoints of (2).

The optimal γ has a closed form. For a single coupling of conditional R^2 equal to ρ^2 estimated from n points, minimizing expected error gives the *reliability*

$$\gamma^* = \frac{(n - 2)\rho^2}{(n - 2)\rho^2 + (1 - \rho^2)}, \quad (3)$$

a Wiener/James–Stein shrinkage [James and Stein, 1961, Ledoit and Wolf, 2012] that tends to 1 as data accrues or coupling strengthens and to 0 when the coupling is unreliable. The shrinkage *intensity* γ^* is classical—an analytic, target-specific shrinkage of the kind Ledoit and Wolf [2004] and Schäfer and Strimmer [2005] derive for covariances and for (partial) correlations. What is particular to the Schur reading is not the

intensity but the *operation* it scales. It is *not* the raw off-diagonal entries shrunk elementwise toward zero; it is the conditional regression b_k and its Schur complement S_k , damped through (2)—a structured, positive-definiteness-preserving shrinkage of the *conditional*. Damping b_k by γ scales the cross-covariance by γ but also adjusts the within-block term (for two unit-variance variables it sends $R \mapsto \begin{bmatrix} 1 & \gamma\rho \\ \gamma\rho & 1-\gamma(1-\gamma)\rho^2 \end{bmatrix}$, not $\begin{bmatrix} 1 & \gamma\rho \\ \gamma\rho & 1 \end{bmatrix}$), so it is genuinely different from the elementwise off-diagonal shrinkage of Schäfer and Strimmer [2005]. That operation is exactly the Vecchia neighbour-conditioning and, read the other way, the hedged residual risk of an allocation. So the quantity that tells a weather model how far to trust a neighbour regression is the quantity that tells a portfolio how far to tilt from risk parity toward minimum variance. We have not found this equivalence stated in either literature: the spatial line treats the damping as a prior to be fitted and the allocation line treats it as a portfolio dial to be chosen, and neither remarks that it is, in closed form, the same reliability γ^* . That identity is the observation of this note; the shrinkage itself is classical.

A candid note on the oversight. The Schur pseudo-likelihood and its closed-form reliability were introduced by the present author in a precursor [Cotton, 2025], which did cite the Vecchia factorization—but only as a computational device for the block-conditional likelihood, without recognizing that the damping (2) is a *regularized* Vecchia conditioning, nor its place in the scalable-Gaussian-process literature, nor that the same reliability is precisely the knob of Schur-complementary allocation. (We have since learned that ShrinkTM [Chakraborty and Katzfuss, 2025] was independently arriving at the conditional shrinkage at the same time, from the spatial side.) That these were one object went unnoticed when the first note was written; surfacing it is the purpose of this one.

4 What each side already solved

Read as one operation, the two literatures are complementary rather than redundant; each supplies what the other lacks (Table 1).

Table 1: The same Schur damping (2), as developed on each side.

	weather / spatial fields	portfolio allocation
S_k means	conditional covariance	residual (hedged) risk
$\gamma = 0$	composite (block) likelihood	hierarchical risk parity
$\gamma = 1$	full Gaussian likelihood	minimum-variance portfolio
how γ is set	fitted (empirical Bayes, ShrinkTM)	closed-form reliability γ^*
scale technique	Vecchia neighbours, $O(p m^2)$	block / cluster structure

The spatial side contributes *neighbour conditioning and ordering*—maxmin orderings and nearest neighbour sets [Guinness, 2018] that make the damped object computable at $p = 10^5$, and an empirical-Bayes machine for *learning* the damping toward a fitted base [Chakraborty and Katzfuss, 2025]. The allocation side contributes the *closed form* (3) and the recognition that the same γ is an investment decision, not only a regularizer. Neither side had both.

The two sides also shrink toward different *targets*, and the asymmetry is not arbitrary: each shrinks toward the structure it can trust. A spatial field has a credible parametric model—a smooth Matérn covariance—so there is real structure to believe in, and ShrinkTM rationally centers its prior there. Financial returns have no such trustworthy parametric covariance; the founding premise of hierarchical risk parity is precisely that estimated cross-correlations are largely noise, so the safe, strong prior is independence ($\rho = 0$, the $\gamma = 0$ / HRP limit). Where weather has structure to believe, finance has structure to doubt—so it is unsurprising, in hindsight, that the spatial road centers on a base GP while the allocation road errs toward zero coupling. The operation (2) and its optimal intensity (3) are the same; only the prior they lean on differs, in the direction each domain’s experience warrants.

5 An example: importing Vecchia into allocation

The clearest way to show the connection is useful is to carry a tool across it. The present author works on the financial side, so the transfer we demonstrate runs in the easier direction—importing the spatial community’s neighbour conditioning and fitted shrinkage into allocation; the reverse import (the closed form and the decision-theoretic reading, into spatial modeling) we can only conjecture.

We apply Vecchia conditioning to daily asset returns—a setting with no parametric base, using a correlation-based neighbour ordering of the kind the spatial literature adopts when Euclidean distance is unavailable—and compare three settings of the damping (2): undamped ($\gamma = 1$, plain Vecchia / minimum-variance conditioning), the closed-form reliability $\gamma^\star$, and a single intensity tuned on a held-out split (the spatial community’s *fit-the-shrinkage* instinct). The external criterion is out-of-sample log-likelihood, swept over the aspect ratio n/p ($p = 60$ assets, $m = 15$ neighbours).

Table 2: Out-of-sample log-likelihood on asset returns (higher is better), by damping. $\bar{\gamma}^\star$ is the mean closed-form reliability; γ_{tuned} the held-out-tuned intensity.

n/p	undamped ($\gamma=1$)	closed-form $\gamma^\star$	tuned γ	$\bar{\gamma}^\star$	γ_{tuned}
0.5	−188.1	−73.7	−37.4	0.65	0.43
0.8	−64.4	−39.0	−28.4	0.71	0.58
1.2	−34.2	−25.0	−19.8	0.77	0.71
2.0	−22.6	−19.4	−17.3	0.84	0.79
3.0	−17.7	−16.1	−15.1	0.88	0.86

Two things follow (Table 2). First, damping the conditioning helps enormously when undersampled: the closed-form $\gamma^\star$ improves on undamped Vecchia by of order 100 nats at $n/p = 0.5$, the gap vanishing by $n/p = 3$ as the conditioning becomes reliable ($\bar{\gamma}^\star \rightarrow 1$). So the spatial idea of regularizing the Vecchia conditional transfers intact to returns, with no parametric base. Second—and this is the spatial community teaching the financial one—a single *fitted* intensity beats the per-point closed form at every ratio: $\gamma^\star$ systematically under-damps ($\bar{\gamma}^\star > \gamma_{\text{tuned}}$ throughout). The lesson the allocation side should take is therefore not the closed form but the *fitting*: treat γ as something to learn, as ShrinkTM does, rather than to read off. The damped estimator is computable in a single $O(p m^2)$ streaming pass that never forms the dense matrix, verified to reproduce the batch estimate to machine precision—so the borrowed tool also fits the online, large- p setting in which allocation actually operates.

6 The reverse import: robustifying the hedge

The transfer also runs the other way. The damping (2) can be used as an *estimator*—form each block’s hedge b_k and Schur complement S_k , damp by $\gamma^\star$, and reassemble the implied positive-definite covariance—rather than by shrinking entries. Applied naively, with $b_k = R_{cc}^{-1} R_{ck}$, this structured estimator is *worse* than elementwise entry shrinkage and blows up when $n \approx p$ (on a grouped recovery task its out-of-sample log-likelihood degrades to -3.5×10^7 at $n/p = 1$), because the structured operation inherits precisely the unstable inverse R_{cc}^{-1} the exercise is trying to avoid. The allocation side’s standard fix for an unstable hedge—a regularized (ridge / shrunk) regression in place of the raw inverse—repairs it. Computing b_k through a Ledoit–Wolf-shrunk conditioning block removes the blow-up and makes the structured estimator the *best* on out-of-sample likelihood, beating both entry shrinkage and a block-aware Ledoit–Wolf at every aspect ratio (e.g. −11.5 vs. −11.8 vs. −20.6 nats at $n/p = 1$); entry shrinkage remains marginally better on raw correlation recovery—a clean match of operation to objective. So the structured Schur damping is viable after all, *given* the allocation side’s robust hedge; it ships as SchurConditionalCovariance alongside

the entry-shrinkage `SchurLedoitWolfCovariance` in `precise`. This is the mirror of Section 5: there the spatial habit of *fitting* the intensity instructed finance; here the financial habit of *regularizing a hedge* rescues the spatial-style structured estimator.

7 Discussion

The point is the correspondence itself. That hierarchical risk parity and the minimum-variance portfolio are the $\gamma = 0$ and $\gamma = 1$ ends of the very damping that a weather model applies to stay estimable is not a metaphor—it is the same Schur complement (1) and the same convex combination (2), read once as a conditional variance and once as a residual risk. Practical consequences run both ways. Allocation can borrow the spatial machinery: neighbour/cluster orderings and empirical-Bayes-fitted damping in place of a fixed γ . Spatial modeling can borrow the allocation reading: the closed-form reliability as a tuning-free initializer, and the reminder that the damping is a decision with a cost, not merely a prior. And both sit on one online primitive—a Schur complement of a few neighbour blocks, damped by a reliability—maintainable incrementally in $O(p m^2)$. This is the form in which the online covariance library `precise`¹ serves both: its block-covariance and Schur–Ledoit–Wolf estimators maintain exactly these damped neighbour-block Schur complements in a single streaming pass, so weather and portfolios run on the same code. The experiment of Section 5 is built on it.

It is fitting that the bridge is a Schur complement. Within finance the same parameter γ already reaches across one divide—from Markowitz’s minimum-variance portfolio at $\gamma = 1$ to hierarchical risk parity at $\gamma = 0$. This note only follows the identical complement across a wider gap, from finance to meteorology, where it had been waiting all along as the conditional variance of a Gaussian field.

References

- Anirban Chakraborty and Matthias Katzfuss. Learning non-gaussian spatial distributions via bayesian transport maps with parametric shrinkage. *arXiv preprint arXiv:2409.19208*, 2025.
- Peter Cotton. Schur complementary portfolios: A unification of machine learning and optimization-based allocation. Geek Culture (Medium), November 2022. <https://medium.com/geekculture/schur-complementary-portfolios-fix-hierarchical-risk-parity-28b0efa1f35f>.
- Peter Cotton. Schur complementary allocation: A unification of hierarchical risk parity and minimum variance portfolios. *arXiv preprint arXiv:2411.05807*, 2024.
- Peter Cotton. Schur pseudo-likelihood: Scoring and regularizing correlation in high dimensions, 2025. Note, `microprediction/precise`.
- Joseph Guinness. Permutation and grouping methods for sharpening gaussian process approximations. *Technometrics*, 60(4):415–429, 2018.
- W. James and Charles Stein. Estimation with quadratic loss. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 361–379, 1961.
- Matthias Katzfuss and Joseph Guinness. A general framework for vecchia approximations of gaussian processes. *Statistical Science*, 36(1):124–141, 2021.

¹<https://github.com/microprediction/precise>

- Matthias Katzfuss and Florian Schäfer. Scalable bayesian transport maps for high-dimensional non-gaussian spatial fields. *Journal of the American Statistical Association*, 119(546):1409–1423, 2024.
- Olivier Ledoit and Michael Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411, 2004.
- Olivier Ledoit and Michael Wolf. Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics*, 40(2):1024–1060, 2012.
- Marcos López de Prado. Building diversified portfolios that outperform out of sample. *The Journal of Portfolio Management*, 42(4):59–69, 2016.
- Juliane Schäfer and Korbinian Strimmer. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical Applications in Genetics and Molecular Biology*, 4(1):Article 32, 2005.
- Aldo V. Vecchia. Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society: Series B*, 50(2):297–312, 1988.