

7547959-less-is-more-forecast-granularity-estimation-error-and-portfolio-choice

SUMMARY: This paper examines the impact of forecast granularity on portfolio choice and performance. The authors derive a pricing model for forecast granularity, where each dimension of the forecast costs $1/T$, the variance of an estimated Sharpe ratio. The paper also considers the impact of estimating the covariance matrix, which adds a Wishart factor that grows with the ratio of dimensions to observations. The authors find that shrinkage pays the price on every dimension, while coarsening pays it only on the dimensions kept and wins when the signal is concentrated. The paper applies these findings to a machine-learning forecast of 100 large CRSP stocks, finding that the optimal number of rank groups is $G^* = 4$ for an information coefficient of 0.05. The authors conclude that the familiar rules (plug-in, Kan-Zhou, Tu-Zhou, minimum variance, $1/N$, decile sorts, parametric tilts) are fixed cells of the resulting grid, and that the optimal number of groups is tied to a measurable property of the forecast.

METHOD: The authors derive a pricing model for forecast granularity, where each dimension of the forecast costs $1/T$, the variance of an estimated Sharpe ratio. They also consider the impact of estimating the covariance matrix, which adds a Wishart factor that grows with the ratio of dimensions to observations. The findings are applied to a machine-learning forecast of 100 large CRSP stocks, with the optimal number of rank groups found to be $G^* = 4$ for an information coefficient of 0.05.

KEY FINDINGS:

- Each dimension of the forecast costs $1/T$, the variance of an estimated Sharpe ratio.
- Coarsening pays the price only on the dimensions kept and wins when the signal is concentrated.
- The optimal number of rank groups is $G^* = 4$ for an information coefficient of 0.05.
- The familiar rules (plug-in, Kan-Zhou, Tu-Zhou, minimum variance, $1/N$, decile sorts, parametric tilts) are fixed cells of the resulting grid.

IMPLICATION FOR TRADING: The findings suggest that investors should consider the granularity of their forecasts when choosing how much of the forecast to use. The optimal number of rank groups is tied to a measurable property of the forecast. The paper recommends using the optimal number of groups, rather than the familiar rules, to maximize performance. The findings have implications for portfolio managers and investors who use machine learning or other types of forecasts.

Less Is More: Forecast Granularity, Estimation Error, and Portfolio Choice[‡]

Lukas Salcher^{‡§}

Sebastian Stöckl^{‡¶}

Michael Hanke^{‡||}

This version: October, 2026

Abstract

An investor who acts on a return forecast chooses how much of it to use. From one identity for the certainty-equivalent loss in risk units we derive the price of every such choice. Each dimension of the forecast the investor acts on costs $1/T$, the variance of an estimated Sharpe ratio, whatever signal it carries; each dimension in which he estimates the covariance matrix adds a Wishart factor that grows with the ratio of dimensions to observations. The familiar rules (plug-in, Kan–Zhou, Tu–Zhou, minimum variance, $1/N$, decile sorts, parametric tilts) are fixed cells of the resulting grid. The cell that moves sets the number of rank groups from the forecast’s information coefficient and holds $1/N$ when the forecast cannot rank. On a hundred large CRSP stocks with a machine-learning forecast it earns an annualized Sharpe ratio of 1.2, against 0.6 for $1/N$ and 1.0 for the decile sort.

Keywords: parameter uncertainty, estimation risk, forecast granularity, risk-adjusted forecast error, portfolio sorts

JEL classification: G11, G17, C53

*Preliminary version. Proofs are given in full in the text of this version; in a journal version they will be moved to an appendix.

[†]We thank Thomas Dangl, Wolfgang Schadner, Roland Mestel, Merlin Bartel, Michael Graff (discussant), Pierre Jaffard (discussant), and Hayeon Park (discussant), as well as participants at the 35th Workshop of the Austrian Working Group on Banking and Finance, the 2021 World Finance Conference, the 26th International Conference on Macroeconomic Analysis and International Finance (ICMAIF), the 2022 European Conference on Stochastic Optimization and Computational Management Science (ECSO-CMS), the 37th Australasian Finance and Banking Conference, and the poster session at the 2026 Annual Meeting of the German Finance Association (DGF), and seminar participants at the University of Neuchâtel and the University of St. Gallen, for helpful comments and suggestions.

[‡]University of Liechtenstein, Liechtenstein Business School, Fürst-Franz-Josef-Strasse, 9490 Vaduz, Liechtenstein.

[§]Email: lukas.salcher@uni.li.

[¶]Corresponding author. Email: sebastian.stoeckl@uni.li.

^{||}Email: michael.hanke@uni.li.

1 Introduction

Mean–variance portfolios built from estimated moments perform poorly out of sample (Jobson and Korkie, 1980; Michaud, 1989; Kan and Zhou, 2007), often worse than $1/N$ (DeMiguel et al., 2009b). The literature offers two kinds of remedy. Academic rules shrink: they keep every dimension of the forecast and the covariance matrix and scale (Jorion, 1986; Ledoit and Wolf, 2004; Kan and Zhou, 2007; Tu and Zhou, 2011). Practitioners coarsen: they sort assets into deciles, tilt on a characteristic or hold equal weights, and act on one or two dimensions of the forecast only (Fama and French, 1993; Brandt et al., 2009). Which is right depends on the forecast, yet each rule fixes its granularity once and for all.

This paper prices granularity. The starting point is an identity: the certainty-equivalent loss of any position is the squared distance between the position and the optimum, measured in risk units (Lemma 1); for the plug-in position it is the squared risk-adjusted forecast error of Salcher et al. (2026). Expected loss then splits into signal discarded and noise retained, and the noise has a price per dimension: every dimension of the forecast the investor acts on costs $1/T$, the variance of an estimated Sharpe ratio, whatever signal it carries (Proposition 3). A dimension is worth keeping if and only if its true Sharpe ratio exceeds its standard error (Corollary 1). Shrinkage pays the price on every dimension; coarsening pays it only on the dimensions kept and wins when the signal is concentrated.

The covariance side has the same structure. An investor who trades K basis portfolios, for instance rank-group portfolios, estimates only their $K \times K$ covariance matrix and pays a Wishart factor that depends on K/T instead of N/T . The two choices nest: the basis decides what is tradable, a reduction of the forecast inside the basis decides how much of it is used. Theorems 1 and 2 give the expected loss of every combination in closed form, with known and with estimated covariance matrix. The familiar rules are fixed cells of this grid (Section 3, Table 1): the plug-in, the Kan–Zhou and Tu–Zhou rules, minimum variance, $1/N$, the decile sort, the parametric tilt and rank tiers. None sets its granularity from what the forecast can deliver.

The cell that moves follows from the same algebra. For a forecast that ranks the assets with information coefficient ρ , the optimal number of rank groups is $G^* = (2NT\rho^2/(1-\rho^2))^{1/3}$ (Proposition 7). The cube root is the familiar rate of a partitioning estimator (Cattaneo et al., 2020), but the price per group is $1/T$ regardless of how many assets the group averages, so G^* is small: about four for a hundred stocks, ten years of data and an information coefficient of 0.05. With $\rho \rightarrow 0$ the rule collapses to one group, i.e. to $1/N$. Section 4 turns this into a six-step rule and applies it to the same hundred large CRSP stocks with six forecasts of increasing quality. In this application the rule never falls below

$1/N$, and its Sharpe ratio rises with forecast quality, from 0.57 to 1.22; the decile sort and the parametric tilt, which fix their granularity, lose on weak forecasts.

The closest predecessor is [Branger et al. \(2019\)](#), who optimize over equally weighted groups of sorted assets and derive an optimal group size. A single group size makes two choices at once, the basis in which the covariance matrix is estimated and the number of forecast dimensions acted on; we separate them, price each exactly and tie the number of groups to a measurable property of the forecast. We extend the expected-loss analysis of [Kan and Zhou \(2007\)](#) from the asset basis to any basis and any fixed reduction of the forecast, and the result does not contradict the virtue of complexity in return prediction ([Kelly et al., 2024](#)): a complex model may well deliver a forecast whose best use is coarse.

Section 2 derives the prices, Section 3 places the familiar rules on the grid, Section 4 states the rule and the illustration, and Section 5 concludes. Proofs are in the text; the appendix collects two technical results.

2 Theory

This section derives, from one identity, what a mean–variance investor loses by acting on an estimated return forecast and an estimated covariance matrix, and how the loss depends on how much of each he uses. The derivation has three parts. First, the welfare loss from any position is the squared distance between the position and the optimum in risk units, and for the plug-in position this is the squared risk-adjusted forecast error (RAFE) of [Salcher et al. \(2026\)](#), the forecast error measured in risk units (Section 2.1). Second, expected loss splits into signal discarded and noise retained, and noise is priced per dimension, so that using fewer dimensions of a forecast can raise welfare (Sections 2.2–2.3). Third, the investor makes two such choices, one for the covariance matrix and one for the forecast, and they nest: he first chooses which portfolios to trade, which fixes the dimension in which the covariance matrix is estimated, and then how much of the forecast to use within that space (Sections 2.4–2.6). The length of the position is a third, separate choice (Section 2.7). Section 3 places the familiar portfolio rules on the grid that these choices span.

2.1 The loss in risk units

There are N risky assets with excess returns r_t , mean μ and positive-definite covariance matrix Σ , and a risk-free asset. A mean–variance investor with risk aversion γ chooses weights $w \in \mathbb{R}^N$ to maximize the certainty equivalent

$$\text{CE}(w) = w' \mu - \frac{\gamma}{2} w' \Sigma w. \tag{1}$$

With μ and Σ known the solution is $w^* = \Sigma^{-1}\mu/\gamma$ and $\text{CE}^* = \mu'\Sigma^{-1}\mu/(2\gamma) = \theta^2/(2\gamma)$, where $\theta = (\mu'\Sigma^{-1}\mu)^{1/2}$ is the maximum Sharpe ratio.

We measure positions and forecasts in *risk units*. With $\Sigma^{1/2}$ the symmetric square root of Σ , let

$$a = \Sigma^{-1/2}\mu, \quad v = \Sigma^{1/2}w. \quad (2)$$

Then $w'\mu = v'a$ and $w'\Sigma w = \|v\|^2$, so $\text{CE}(w) = v'a - \frac{\gamma}{2}\|v\|^2$. The vector a lists the Sharpe ratios of N orthogonal directions, $\|a\| = \theta$, and the optimum is $v^* = a/\gamma$.

Lemma 1 (Loss identity). *For any position w with $v = \Sigma^{1/2}w$,*

$$\text{CE}^* - \text{CE}(w) = \frac{1}{2\gamma} \|\gamma v - a\|^2. \quad (3)$$

Proof. In risk units, $\text{CE}^* = \|a\|^2/(2\gamma)$ and $\text{CE}(w) = v'a - \frac{\gamma}{2}\|v\|^2$. Subtract and put everything over the common factor $1/(2\gamma)$:

$$\text{CE}^* - \text{CE}(w) = \frac{1}{2\gamma} \|a\|^2 - v'a + \frac{\gamma}{2} \|v\|^2 = \frac{1}{2\gamma} (\|a\|^2 - 2\gamma v'a + \gamma^2 \|v\|^2).$$

The bracket is the expansion of $\|\gamma v - a\|^2 = \gamma^2 \|v\|^2 - 2\gamma v'a + \|a\|^2$. □

Suppose the investor holds a forecast $\hat{\mu}$ of μ and, for now, knows Σ . The plug-in position $w = \Sigma^{-1}\hat{\mu}/\gamma$ has $\gamma v = \hat{b}$, where

$$\hat{b} = \Sigma^{-1/2}\hat{\mu} \quad (4)$$

is the forecast in risk units, and Lemma 1 gives

$$\text{CE}^* - \text{CE}(w) = \frac{1}{2\gamma} \|\hat{b} - a\|^2. \quad (5)$$

The distance $\|\hat{b} - a\|$ is the risk-adjusted forecast error (RAFE) of [Salcher et al. \(2026\)](#), who write $m = \Sigma^{-1/2}\mu$ and $\hat{m} = \Sigma^{-1/2}\hat{\mu}$ for our a and $\hat{b}$ and bound the Sharpe-ratio gap of the plug-in portfolio by a multiple of it; we keep a and $\hat{b}$ because m is needed below for the means of the basis portfolios, and because (5) is an identity for the certainty-equivalent loss rather than a bound for the Sharpe ratio (Section 2.7 connects the two). The identity is the loss function of the estimation-risk literature ([Klein and Bawa, 1976](#); [Kan and Zhou, 2007](#)) written in risk units; what is used below is that it holds for every position, not only for the plug-in. Equation (5) is exact for every forecast, whatever its origin: welfare is lost only through forecast error, and forecast error matters only in risk units. Everything below decomposes $\|\hat{b} - a\|^2$, or its expectation, for positions that use less than all of $\hat{b}$.

2.2 Forecast error and the price of a dimension

Write the forecast in risk units as truth plus error,

$$\hat{b} = a + \eta, \quad \mathbb{E}[\eta] = 0, \quad \text{Var}(\eta) = \Psi. \quad (6)$$

Only the first two moments of η are used until Section 2.4.

Proposition 1 (Expected loss of the plug-in). *Under (6), $2\gamma \mathbb{E}[\text{CE}^* - \text{CE}(w)] = \text{tr } \Psi$.*

Proof. By (5), $2\gamma \text{loss} = \|\hat{b} - a\|^2 = \|\eta\|^2 = \eta' \eta$. For any random vector with mean zero, $\mathbb{E}[\eta' \eta] = \sum_i \mathbb{E}[\eta_i^2] = \sum_i \text{Var}(\eta_i) = \text{tr } \Psi$. $\square$

The leading case is the historical mean. If $\hat{\mu}$ is the sample mean of T independent monthly observations, $\text{Var}(\hat{\mu}) = \Sigma/T$ and therefore $\Psi = I_N/T$: in risk units the error is isotropic, and

$$2\gamma \mathbb{E}[\text{loss}] = \frac{N}{T}. \quad (7)$$

Each of the N dimensions of the forecast costs $1/T$, however much or little signal it carries. We write

$$s^2 = \frac{1}{T} \quad (8)$$

for this *price of a dimension*: the variance of an estimated Sharpe ratio when the volatility is known, the same for every direction; with an estimated volatility it is $(1 + \text{SR}^2/2)/T$ (Jobson and Korkie, 1981; Lo, 2002), which differs from $1/T$ by a few percent at most for monthly Sharpe ratios. With ten years of monthly data $s = 0.09$ per month; a hundred assets add 0.83 to the squared Sharpe ratio the investor believes he has, none of which exists. Equation (7) is the estimation-risk result of Kan and Zhou (2007) for known Σ ; writing it as $\text{tr } \Psi$ shows that it holds for any forecast with error covariance Ψ .

2.3 Using less of a forecast: soft and hard

Since every retained dimension costs s^2 , the investor may gain by not acting on all of $\hat{b}$. There are two ways to use less of a vector: shrink every component (soft), or set some components to zero and keep the others unchanged (hard).

Soft. The position is $\gamma v = \kappa \hat{b}$, $\kappa \in [0, 1]$.

Proposition 2 (Soft reduction). *Under (6), $2\gamma \mathbb{E}[\text{loss}](\kappa) = (1 - \kappa)^2 \|a\|^2 + \kappa^2 \text{tr } \Psi$, minimized at*

$$\kappa^* = \frac{\|a\|^2}{\|a\|^2 + \text{tr } \Psi}, \quad 2\gamma \mathbb{E}[\text{loss}](\kappa^*) = \frac{\|a\|^2 \text{tr } \Psi}{\|a\|^2 + \text{tr } \Psi}. \quad (9)$$

Proof. Insert $\hat{b} = a + \eta$ into the loss identity with $\gamma v = \kappa \hat{b}$:

$$\kappa \hat{b} - a = \kappa a + \kappa \eta - a = (\kappa - 1)a + \kappa \eta.$$

Expand the squared norm:

$$\|(\kappa - 1)a + \kappa \eta\|^2 = (1 - \kappa)^2 \|a\|^2 + 2\kappa(\kappa - 1) a' \eta + \kappa^2 \|\eta\|^2.$$

Take expectations. The middle term vanishes because $\mathbb{E}[\eta] = 0$ and a is a constant; the last is $\kappa^2 \text{tr } \Psi$ by the argument of Proposition 1. This gives the loss as a function of κ . It is a convex quadratic in κ ; setting its derivative $-2(1 - \kappa)\|a\|^2 + 2\kappa \text{tr } \Psi$ to zero gives $\kappa^*(\|a\|^2 + \text{tr } \Psi) = \|a\|^2$. Substituting κ^* back, $1 - \kappa^* = \text{tr } \Psi / (\|a\|^2 + \text{tr } \Psi)$, and

$$(1 - \kappa^*)^2 \|a\|^2 + (\kappa^*)^2 \text{tr } \Psi = \frac{(\text{tr } \Psi)^2 \|a\|^2 + \|a\|^4 \text{tr } \Psi}{(\|a\|^2 + \text{tr } \Psi)^2} = \frac{\|a\|^2 \text{tr } \Psi (\text{tr } \Psi + \|a\|^2)}{(\|a\|^2 + \text{tr } \Psi)^2},$$

which cancels to the stated minimum. $\square$

Proposition 2 is the oracle form of James–Stein shrinkage (James and Stein, 1961) toward zero; in portfolio choice it underlies the Bayes–Stein estimator of Jorion (1986) and the two-fund rule of Kan and Zhou (2007). For the historical mean, $\kappa^* = \theta^2 / (\theta^2 + N/T)$, the two-fund rule of Kan and Zhou (2007) for known Σ . The forecast’s quality does not appear separately because for this forecast it is not free: $\Psi = I/T$ is fixed by the window. Shrinkage pays on every dimension: a fraction $(1 - \kappa)^2$ of its signal and $\kappa^2 s^2$ of noise, whether or not the dimension carries any signal.

Hard. The position is $\gamma v = P\hat{b}$ with P an orthogonal projection ($P = P' = P^2$) of rank $k \leq N$: the investor acts on the k dimensions of the forecast in the range of P and ignores the rest.

Proposition 3 (Hard reduction). *For any orthogonal projection P of rank k and every realization of η ,*

$$2\gamma \text{loss} = \|P\hat{b} - a\|^2 = \|(I - P)a\|^2 + \|P\eta\|^2, \quad (10)$$

so that

$$2\gamma \mathbb{E}[\text{loss}] = \underbrace{\|(I - P)a\|^2}_{\text{signal discarded}} + \underbrace{\text{tr}(P\Psi)}_{\text{noise retained}} = \|(I - P)a\|^2 + k s^2 \quad \text{if } \Psi = s^2 I. \quad (11)$$

Proof. Insert $\hat{b} = a + \eta$ and split a into the part inside the range of P and the part outside, $a = Pa + (I - P)a$:

$$P\hat{b} - a = Pa + P\eta - Pa - (I - P)a = -(I - P)a + P\eta.$$

Expand the squared norm:

$$\|P\hat{b} - a\|^2 = \|(I - P)a\|^2 - 2a'(I - P)'P\eta + \|P\eta\|^2.$$

The cross term contains $(I - P)'P = (I - P)P = P - P^2 = 0$, using $P' = P$ and $P^2 = P$. So it is zero for every realization of η , not only on average, which proves (10). For the expectation, write $\|P\eta\|^2 = \eta'P'P\eta = \eta'P\eta$ and use the trace trick $\mathbb{E}[\eta'P\eta] = \text{tr}(P\mathbb{E}[\eta\eta']) = \text{tr}(P\Psi)$. If $\Psi = s^2I$ this is $s^2 \text{tr} P$, and the trace of an orthogonal projection equals its rank k , because its eigenvalues are k ones and $N - k$ zeros. $\square$

Equation (11) is a bias–variance decomposition in risk units: the squared Sharpe ratio of the part of the truth the investor no longer sees, plus the number of dimensions he still acts on times the price of a dimension. Compare the two reductions. In the soft case the noise carried is $\kappa^2 \text{tr} \Psi$, and since κ^* rises with $\|a\|^2$, how much noise the investor ends up holding is decided by how much signal there is. In the hard case the noise term is fixed by the count k alone: whether the retained directions carry much signal or little, each costs s^2 . The signal enters only through the first term, and there only the part that is dropped. The decomposition is the bias–variance trade-off of a projection estimator (Hastie et al., 2009, Ch. 3), and Kozak et al. (2020) use the same trade-off when they truncate the principal components of the stochastic discount factor. What the certainty-equivalent loss adds is the unit: every retained dimension costs s^2 of squared Sharpe ratio, whatever it carries.

Proposition 3 charges $\text{tr}(P\Psi)$ for any k directions; the next step asks which k , and the answer is a test applied to one direction at a time.

Corollary 1 (Keep or kill). *Let $u_1, \dots, u_N$ be an orthonormal basis and $P_K = \sum_{j \in K} u_j u_j'$ the projection onto the directions in a set K .*

(a) *For any error covariance Ψ ,*

$$2\gamma \mathbb{E}[\text{loss}](K) = \sum_{j \notin K} (u_j' a)^2 + \sum_{j \in K} u_j' \Psi u_j, \quad (12)$$

minimized by $K^ = \{j : (u_j' a)^2 > u_j' \Psi u_j\}$: keep a direction if its squared Sharpe ratio exceeds the variance of its own forecast error.*

(b) *If the error is isotropic, $\Psi = s^2 I$, every direction has the same hurdle and*

$$2\gamma \mathbb{E}[\text{loss}](K) = \sum_{j \notin K} (u_j' a)^2 + |K| s^2, \quad K^* = \{j : (u_j' a)^2 > s^2\}. \quad (13)$$

Proof. P_K is an orthogonal projection of rank $|K|$, so Proposition 3 applies. Expand a in the basis, $a = \sum_j (u'_j a) u_j$; then $(I - P_K)a = \sum_{j \notin K} (u'_j a) u_j$ and, by orthonormality, $\|(I - P_K)a\|^2 = \sum_{j \notin K} (u'_j a)^2$. For the noise term, $\text{tr}(P_K \Psi) = \sum_{j \in K} \text{tr}(u_j u'_j \Psi) = \sum_{j \in K} u'_j \Psi u_j$. The loss is therefore a sum over directions in which direction j contributes $(u'_j a)^2$ if it is dropped and $u'_j \Psi u_j$ if it is kept, and each direction can be decided separately. With $\Psi = s^2 I$, $u'_j \Psi u_j = s^2 \|u_j\|^2 = s^2$. $\square$

Under isotropy a direction is worth acting on if and only if its true Sharpe ratio exceeds s , the standard error of an estimated Sharpe ratio. In statistical terms this is oracle hard thresholding in the basis $u_1, \dots, u_N$ (Hastie et al., 2009, Ch. 3); the threshold is the standard error of an estimated Sharpe ratio. Without isotropy the test is the same but the hurdle is the direction's own noise: a direction with much signal but a noisy forecast is dropped, one with little signal but a precise forecast is kept.

Remark 1 (When the error is isotropic). For the sample mean, isotropy is not an assumption but a consequence of measuring in risk units: $\text{Var}(\hat{\mu}) = \Sigma/T$ has the covariance structure of the returns, and whitening by $\Sigma^{-1/2}$ removes it exactly, $\Psi = I/T$. For a general forecast with error covariance Ω in return units, $\Psi = \Sigma^{-1/2} \Omega \Sigma^{-1/2}$, which is isotropic if and only if $\Omega \propto \Sigma$. A forecast built from the same returns (a moving average, a regression on factors) inherits the structure of Σ approximately; an exogenous forecast (a characteristic, a machine-learning score) need not. The polar case is an error that is independent across assets with a common variance, $\Omega = \omega^2 I$: then $\Psi = \omega^2 \Sigma^{-1}$, and the noise sits in the low-variance directions of Σ , exactly where $\Sigma^{-1} \mu$ places its largest weights. Two consequences follow when Ψ is not isotropic. The ordering of candidate directions changes: with $\Psi = s^2 I$ every orthonormal basis carries the same noise and the basis matters only through how it concentrates the signal; in general the hurdle of a direction is its own noise $u' \Psi u$, and a basis is good if it concentrates the signal on directions whose forecast error is small. (Over all projections the oracle keeps at most one direction, the leading eigenvector of $aa' - \Psi$, since $2\gamma \mathbb{E}[\text{loss}] = \|a\|^2 - \text{tr}(P(aa' - \Psi))$ and $aa' - \Psi$ has at most one positive eigenvalue; this is an oracle statement and not a rule.) And the noise retained by a group dimension is $1'_g \Psi 1_g / n_g$, which equals s^2 under isotropy and exceeds it when errors within a group are positively correlated; Corollary 3 below assumes isotropy, and its G^* falls when within-group errors are positively correlated. Whether a given forecast's error is close to isotropic is an empirical question, answered by the covariance of the cross-sectional residuals of returns on the forecast relative to Σ ; Proposition 7 states the assumptions under which the rule's inputs are observable, and the application checks them for the forecasts used in this paper.

Remark 2 (Soft within hard). $\gamma v = \kappa P \hat{b}$ has expected loss $\|(I - P)a\|^2 + (1 - \kappa)^2 \|Pa\|^2 + \kappa^2 ks^2$, minimized at $\kappa_P^* = \|Pa\|^2 / (\|Pa\|^2 + ks^2)$: every hard cell has its own scalar, with k in place of N .

Remark 3 (When hard beats soft). If the signal is concentrated on k_0 directions, $(u'_j a)^2 = \|a\|^2/k_0$ for $j \in K_0$ and zero otherwise, hard reduction with $K = K_0$ loses $k_0 s^2$, while soft reduction at its best loses $\|a\|^2 N s^2 / (\|a\|^2 + N s^2)$. The soft loss is the larger one exactly when the signal per retained direction exceeds $s^2 N / (N - k_0)$, i.e. $\|a\|^2/k_0 > s^2 N / (N - k_0)$, which for $k_0 \ll N$ is the keep-or-kill threshold itself. Soft reduction cannot exploit concentration because it treats every dimension alike.

Two things about this section should be kept in mind. A “direction” is so far any unit vector u in the space of risk units, and P is the orthogonal projection onto any k -dimensional subspace: eigenvectors of Σ , factor exposures, single assets after whitening, or the group indicators of a ranking, which are the subject of Section 2.6. Proposition 3 says that keeping k directions costs ks^2 whichever they are; which k to keep is a separate question. And the projection is not yet something the investor can execute: a direction u in risk units is the portfolio $\Sigma^{-1/2}u$ in asset space, which he cannot form without knowing Σ . Section 2.4 shows that restricting the portfolios he trades is itself a projection in risk units, onto a subspace fixed by his choice of portfolios, and that it requires only their $K \times K$ covariance matrix.

2.4 The action space: which portfolios to trade

Suppose the investor restricts the set of portfolios he trades. With Σ known this is one more way to use less of the forecast; its own rationale, developed below, is that it reduces the covariance matrix he has to estimate. Let D be an $N \times K$ matrix of full column rank whose columns are $K \leq N$ portfolios, the *basis portfolios*, and let the investor choose $w = Dx$, $x \in \mathbb{R}^K$. The basis is his choice: the assets themselves ($D = I$), a single portfolio, a set of factor-mimicking portfolios, principal components, or portfolios formed from a ranking. This section treats D as given and studies what the restriction costs; which D to choose is discussed at the end of the section.

The returns of the basis portfolios are $y_t = D' r_t$ with mean $m = D'\mu$ and covariance $V = D'\Sigma D$, a $K \times K$ matrix. Write

$$\tilde{a} = V^{-1/2}m, \quad \tilde{v} = V^{1/2}x \quad (14)$$

for the truth and the position in risk units *within* the action space. Then $\text{CE}(Dx) = \tilde{v}'\tilde{a} - \frac{\gamma}{2}\|\tilde{v}\|^2$, the best position in the space is $\tilde{v} = \tilde{a}/\gamma$, and the best certainty equivalent attainable is $\|\tilde{a}\|^2/(2\gamma)$.

Lemma 2 (Tradable signal). $\|\tilde{a}\|^2 = \|\Pi_D a\|^2$, where $\Pi_D = E(E'E)^{-1}E'$, $E = \Sigma^{1/2}D$, is the orthogonal projection onto $\text{span}(\Sigma^{1/2}D)$. Hence, for any x ,

$$2\gamma[\text{CE}^* - \text{CE}(Dx)] = (\|a\|^2 - \|\Pi_D a\|^2) + \|\gamma\tilde{v} - \tilde{a}\|^2. \quad (15)$$

Proof. First the identity. By definition $\|\tilde{a}\|^2 = m'V^{-1}m$ with $m = D'\mu$ and $V = D'\Sigma D$. Write $\mu = \Sigma^{1/2}a$ and $\Sigma = \Sigma^{1/2}\Sigma^{1/2}$, so that $D'\mu = D'\Sigma^{1/2}a = E'a$ and $D'\Sigma D = E'E$ with $E = \Sigma^{1/2}D$. Then

$$m'V^{-1}m = a'E(E'E)^{-1}E'a = a'\Pi_D a = \|\Pi_D a\|^2,$$

the last step because Π_D is symmetric and idempotent. Π_D is the orthogonal projection onto the column space of E , i.e. onto $\text{span}(\Sigma^{1/2}D)$, which is the set of positions $v = \Sigma^{1/2}w$ that the investor can reach with $w = Dx$.

Now the decomposition. The investor's problem inside the space is a K -dimensional mean-variance problem with mean m and covariance V , so Lemma 1 applies to it with $\tilde{a}$ and $\tilde{v}$ in place of a and v : the gap between the best certainty equivalent in the space, $\|\tilde{a}\|^2/(2\gamma)$, and $\text{CE}(Dx)$ is $\|\gamma\tilde{v} - \tilde{a}\|^2/(2\gamma)$. Add the gap between the unrestricted optimum $\|a\|^2/(2\gamma)$ and the best in the space, $\|\tilde{a}\|^2/(2\gamma) = \|\Pi_D a\|^2/(2\gamma)$, and multiply by 2γ . $\square$

The first term of (15) is the squared Sharpe ratio of the part of the truth that no combination of the basis portfolios can reach: the *signal not tradable*. It is the price of restricting the action space, paid whatever the investor does inside it. $\|\tilde{a}\|^2 = m'V^{-1}m$ is the familiar maximum squared Sharpe ratio attainable with the basis portfolios; the lemma writes its gap to θ^2 as a projection, which is what makes the action space comparable to a reduction of the forecast. The second term is the loss inside the space, and it has exactly the structure of Sections 2.1–2.3 with K in place of N : an investor who uses the forecast's implication for the basis portfolios, $\hat{m} = D'\hat{\mu}$, has $\gamma\tilde{v} = \tilde{b} = V^{-1/2}\hat{m} = \tilde{a} + \tilde{\eta}$, and with $\tilde{\Psi} = \text{Var}(\tilde{\eta})$,

$$2\gamma \mathbb{E}[\text{loss}] = (\|a\|^2 - \|\Pi_D a\|^2) + \text{tr } \tilde{\Psi} = (\|a\|^2 - \|\Pi_D a\|^2) + Ks^2 \quad (16)$$

for the sample mean, since $\text{Var}(\hat{m}) = V/T$ gives $\tilde{\Psi} = I_K/T$. Restricting the action space is therefore a hard reduction: it discards the untradable signal and keeps K dimensions of noise. In the metric of risk units it is the projection Π_D , and Proposition 3 applies with $P = \Pi_D$.

Estimated covariance. The reason to restrict the action space is the covariance matrix. The investor observes the K series y_t and estimates their mean $\hat{m}$ and their $K \times K$ covariance matrix $\hat{V}$; he never estimates a covariance between assets that is not implied

by these. $\hat{V}$ has $K(K+1)/2$ parameters instead of $N(N+1)/2$. His position is

$$w = Dx, \quad x = \frac{1}{\gamma} \hat{V}^{-1} \hat{m}. \quad (17)$$

Assumption 1. $r_1, \dots, r_T$ are i.i.d. $N(\mu, \Sigma)$; $\hat{m}$ and $\hat{V}$ are the sample mean and the unbiased sample covariance matrix of $y_t = D'r_t$; $T > K + 4$. The basis D , and in Section 2.6 the reduction Q , are fixed matrices chosen independently of $r_1, \dots, r_T$.

The last sentence matters: a basis built from the window's own returns (eigenvectors of $\hat{\Sigma}$, groups formed from a sample mean) is not covered, and Remark 8 returns to this.

Remark 4 (Beyond normality). Assumption 1 is used for three things, and it is worth separating them. First, $\text{Var}(\hat{m}) = V/T$: this needs only uncorrelated months with finite variances. Second, the independence of $\hat{m}$ and $\hat{V}$: this is a property of the normal distribution; under skewness the sample mean and the sample covariance matrix are correlated and $\mathbb{E}[\hat{V}^{-1}\hat{m}] \neq \mathbb{E}[\hat{V}^{-1}]\mathbb{E}[\hat{m}]$. Third, the moments (18) of the inverse Wishart matrix. Everything up to this point (Lemma 1 to Lemma 2, and below Theorem 1, Proposition 6 and Corollary 3) uses only second moments and is distribution-free. The results that estimate the covariance matrix (Proposition 4, Proposition 5, Theorem 2, Corollaries 2 and 4, and the scalar of Section 2.7) keep their structure under weaker assumptions, four terms with the covariance price indexed by K and T and the noise count by k , but their constants change. For elliptical returns $\mathbb{E}[\hat{V}^{-1}]$ is still proportional to V^{-1} by symmetry, with constants that grow with the kurtosis, so heavier tails raise α_K and β_K and move the boundary of Remark 6 to a smaller K/T . Under volatility clustering the effective number of independent months is below T , and s^2 and the constants are those of that smaller number. In both cases the covariance price grows faster in K/T than the normal constants say, and the optimal K is smaller. The ordering of the cells in Section 3 by K and k , and the direction of every comparative-static statement, do not depend on the constants; the third decimal of K^* does. The empirical part of the paper does not use the constants: its inference is overlap-robust and heteroskedasticity-consistent, and what it tests is the ordering. Salcher et al. (2026) compare the same loss decomposition under simulated normal returns and on empirical data and find the same qualitative pattern with larger covariance errors in the data.

Under Assumption 1, $\hat{m}$ and $\hat{V}$ are independent, $(T-1)\hat{V}$ is Wishart, and (Kan and Zhou, 2007, Lemma 1; Haff, 1979)

$$\begin{aligned} \mathbb{E}[\hat{V}^{-1}] &= \alpha_K V^{-1}, & \mathbb{E}[\hat{V}^{-1} V \hat{V}^{-1}] &= \beta_K V^{-1}, \\ \alpha_K &= \frac{T-1}{T-K-2}, & \beta_K &= \frac{(T-1)^2(T-2)}{(T-K-1)(T-K-2)(T-K-4)}. \end{aligned} \quad (18)$$

Both constants exceed one, $\beta_K > \alpha_K^2$, and both increase without bound as $K/T \rightarrow 1$. At $T = 120$: $\alpha_5 = 1.05$, $\beta_5 = 1.17$; $\alpha_{30} = 1.35$, $\beta_{30} = 2.48$; $\alpha_{100} = 6.6$, $\beta_{100} \approx 305$.

Proposition 4 (Loss with an estimated covariance matrix). *Under Assumption 1, the position (17) has*

$$2\gamma \mathbb{E}[\text{loss}] = \underbrace{\|a\|^2 - \|\Pi_D a\|^2}_{\text{signal not tradable}} + \underbrace{(1 - 2\alpha_K + \beta_K) \|\Pi_D a\|^2}_{\text{inversion error}} + \underbrace{\beta_K K s^2}_{\text{noise, amplified}}. \quad (19)$$

Proof. The certainty equivalent of $w = Dx$ is $\text{CE}(Dx) = x'm - \frac{\gamma}{2} x'Vx$ with $x = \hat{V}^{-1}\hat{m}/\gamma$. We need the expectations of the two terms.

Mean term. $x'm = \hat{m}'\hat{V}^{-1}m/\gamma$. Because $\hat{m}$ and $\hat{V}$ are independent, take the expectation over $\hat{V}$ first: $\mathbb{E}[\hat{V}^{-1}] = \alpha_K V^{-1}$, then over $\hat{m}$: $\mathbb{E}[\hat{m}] = m$. So $\mathbb{E}[x'm] = \alpha_K m'V^{-1}m/\gamma$.

Variance term. $x'Vx = \hat{m}'\hat{V}^{-1}V\hat{V}^{-1}\hat{m}/\gamma^2$. Over $\hat{V}$ first: $\mathbb{E}[\hat{V}^{-1}V\hat{V}^{-1}] = \beta_K V^{-1}$, so $\mathbb{E}[x'Vx] = \beta_K \mathbb{E}[\hat{m}'V^{-1}\hat{m}]/\gamma^2$. For the remaining expectation write $\hat{m} = m + (\hat{m} - m)$ with $\mathbb{E}[\hat{m} - m] = 0$ and $\text{Var}(\hat{m}) = V/T$:

$$\begin{aligned} \mathbb{E}[\hat{m}'V^{-1}\hat{m}] &= m'V^{-1}m + \mathbb{E}[(\hat{m} - m)'V^{-1}(\hat{m} - m)] \\ &= m'V^{-1}m + \text{tr}(V^{-1}V/T) = m'V^{-1}m + K/T. \end{aligned}$$

Collect. With $m'V^{-1}m = \|\Pi_D a\|^2$ from Lemma 2,

$$2\gamma \mathbb{E}[\text{CE}(Dx)] = 2\alpha_K \|\Pi_D a\|^2 - \beta_K \|\Pi_D a\|^2 - \beta_K K/T,$$

and subtracting from $2\gamma \text{CE}^* = \|a\|^2$:

$$2\gamma \mathbb{E}[\text{loss}] = \|a\|^2 - 2\alpha_K \|\Pi_D a\|^2 + \beta_K \|\Pi_D a\|^2 + \beta_K K s^2.$$

Adding and subtracting $\|\Pi_D a\|^2$ gives the three terms of (19). The moment results for $\hat{V}^{-1}$ are those of Kan and Zhou (2007), Lemma 1, applied to the K portfolio return series. $\square$

The three terms are the price of the covariance side. The first is the signal given up by restricting the space, a hard reduction along risk directions. The second is the cost of inverting a noisy matrix: since $1 - 2\alpha_K + \beta_K = (\alpha_K - 1)^2 + (\beta_K - \alpha_K^2)$ it collects a leverage bias ($\hat{V}^{-1}$ too large on average) and a direction error ($\hat{V}^{-1}$ rotates the position); it is zero when V is known and depends on K and T only, to first order on K/T . The third is the noise term of Section 2.2, now multiplied by β_K . With $K = N$ and $D = I$, (19) is the expected loss of the plug-in portfolio of Kan and Zhou (2007). For K small relative to T

the inversion error is about $(K+1)/T$ times the tradable signal and $\beta_K \approx 1 + 3(K+1)/T$, so (19) exceeds (16) by terms of relative order K/T .

Remark 5 (The covariance error as a matrix: C-RAFE). Proposition 4 is the expectation of a pathwise identity in which the covariance error appears as a matrix. Define

$$\hat{C} = V^{1/2} \hat{V}^{-1} V^{1/2}, \quad (20)$$

the estimated precision matrix of the basis portfolios measured in their true risk units; $\hat{C} = I$ when $\hat{V} = V$. With $x = \hat{V}^{-1} \hat{m} / \gamma$, the position in risk units is $\gamma \tilde{v} = V^{1/2} \hat{V}^{-1} \hat{m} = \hat{C} \tilde{b}$, so by Lemma 2

$$2\gamma \text{loss} = (\|a\|^2 - \|\Pi_D a\|^2) + \|\hat{C} \tilde{b} - \tilde{a}\|^2 \quad (21)$$

for every realization. For $D = I$ this is the quantity that Salcher et al. (2026) study: their precision-error matrix $\Sigma^{1/2} \hat{\Sigma}^{-1} \Sigma^{1/2}$ is our $\hat{C}$, and they call $\|\hat{C} - I\|$ in the spectral norm the covariance RAFE (C-RAFE) and bound the Sharpe-ratio gap by $c_0 \cdot \text{RAFE} + \theta \cdot \text{C-RAFE}$ with $c_0 \in \{1, 2\}$, a worst-case bound that holds path by path. Proposition 4 is the average-case version of the same object: under Assumption 1, $\mathbb{E}[\hat{C}] = \alpha_K I$ and $\mathbb{E}[\hat{C}^2] = \beta_K I$ by (18), so for any fixed unit vector u

$$\mathbb{E}\|(\hat{C} - I)u\|^2 = u' \mathbb{E}[\hat{C}^2 - 2\hat{C} + I] u = \beta_K - 2\alpha_K + 1. \quad (22)$$

The inversion-error term of (19) is therefore the expected squared C-RAFE applied to the tradable signal, $\mathbb{E}\|(\hat{C} - I)\tilde{a}\|^2 = (1 - 2\alpha_K + \beta_K)\|\tilde{a}\|^2$, and the amplification β_K of the noise term is $\mathbb{E}\|\hat{C}\tilde{\eta}\|^2 / \mathbb{E}\|\tilde{\eta}\|^2$. Where the bound uses the largest eigenvalue of $\hat{C} - I$, the expectation uses its second moment; both are functions of K and T only, to first order of K/T .

Which portfolios. Proposition 4 treats D as given, but it also says how to choose it, because every term in (19) is a function of two quantities: the tradable signal $S_K = \|\Pi_D a\|^2$ and the dimension K . Collecting terms,

$$2\gamma \mathbb{E}[\text{loss}] = \|a\|^2 - \varphi_K S_K + \beta_K K s^2, \quad \varphi_K = 2\alpha_K - \beta_K, \quad (23)$$

where φ_K is the fraction of the tradable signal that survives the inversion for the position at unit length: it equals one when V is known, falls with K/T ($\varphi_5 = 0.94$, $\varphi_{30} = 0.22$ at $T = 120$) and reaches zero at $K = 34$, $K/T \approx 0.28$, beyond which, at unit length, no tradable signal lowers the loss. Adding a portfolio to the basis therefore has a benefit and a price, and the comparison is the covariance-side version of Corollary 1.

Proposition 5 (Keep or kill on the covariance side). *Let $D_K = [D_{K-1}, d]$ extend a basis by one portfolio, and let $\Delta_K = S_K - S_{K-1} \geq 0$ be the tradable signal the new portfolio*

adds. Under Assumption 1, adding d lowers the expected loss if and only if

$$\varphi_K \Delta_K > (\varphi_{K-1} - \varphi_K) S_{K-1} + (K\beta_K - (K-1)\beta_{K-1}) s^2. \quad (24)$$

For $K/T \rightarrow 0$ the right-hand side tends to s^2 and the criterion is $\Delta_K > s^2$; it rises with K/T through both terms, the first because the new dimension degrades the inversion of the signal already tradable, the second because it adds a dimension of noise at the amplified price.

Proof. Write $L_K = \|a\|^2 - \varphi_K S_K + \beta_K K s^2$ for the loss with the basis D_K and L_{K-1} for the basis D_{K-1} . Adding the portfolio lowers the loss if $L_K < L_{K-1}$, i.e. if

$$\varphi_K S_K - \beta_K K s^2 > \varphi_{K-1} S_{K-1} - \beta_{K-1} (K-1) s^2.$$

Substitute $S_K = S_{K-1} + \Delta_K$ on the left and move all terms in S_{K-1} and s^2 to the right:

$$\varphi_K \Delta_K > (\varphi_{K-1} - \varphi_K) S_{K-1} + (K\beta_K - (K-1)\beta_{K-1}) s^2.$$

Both bracketed terms are nonnegative: φ_K is decreasing in K and $K\beta_K$ is increasing in K . As $K/T \rightarrow 0$, $\varphi_K \rightarrow 1$, $\varphi_{K-1} - \varphi_K \rightarrow 0$ and $K\beta_K - (K-1)\beta_{K-1} \rightarrow 1$, so the criterion becomes $\Delta_K > s^2$. $\Delta_K \geq 0$ because enlarging a subspace cannot shrink the projection of a onto it. $\square$

A basis portfolio is worth adding if the squared Sharpe ratio it makes newly tradable exceeds the price of estimating one more dimension, and that price starts at s^2 and grows without bound. The criterion ranks candidate portfolios by tradable signal per dimension, and two orderings supply candidates. One is by risk: with $\Sigma = \sum_j \lambda_j u_j u_j'$ the optimum $\Sigma^{-1} \mu = \sum_j (u_j' \mu / \lambda_j) u_j$ weights each eigen-direction by the inverse of its variance, so the low-variance directions can carry much of $\|a\|^2$ per unit of risk while being the ones whose eigenvalues are estimated with the largest relative error; a basis of eigenvectors of Σ (in practice estimated on a separate sample, since eigenvectors of $\hat{\Sigma}$ from the same window violate Assumption 1), ordered by variance, trades the well-estimated directions first. A basis of industry or cluster portfolios is a coarser version of the same ordering (Lopez de Prado, 2016). The other ordering is by forecast: if a forecast can rank the assets, portfolios formed from that ranking concentrate the signal the forecast knows about into few dimensions, which is what (24) rewards; Section 2.6 takes this ordering up. The two orderings need not agree, and a mixed basis, a few risk portfolios together with a few ranking portfolios, is admissible: D is simply their union, and (24) prices each addition the same way.

Remark 6 (Large cross-sections). If $N \geq T$ the sample covariance matrix of the assets is singular and the cell $D = I$ does not exist. The constants (18) require $T > K + 4$, so the action space has a hard ceiling $K \leq T - 5$ whatever the signal, and an effective one far below it: at unit length, a position that inverts a K -dimensional sample covariance matrix loses more than the entire tradable signal once $\beta_K > 2\alpha_K$, asymptotically $K/T > 1 - 1/\sqrt{2} \approx 0.29$ ($K = 34$ at $T = 120$), and is then worse than the risk-free asset even without forecast noise. For $D = I$ this is a property of the expected-loss formula of Kan and Zhou (2007); the threshold $\varphi_K < 0$ states it for any basis of K portfolios. With the optimal length of Section 2.7 the expected certainty equivalent stays positive, $\alpha_K^2 S_K^2 / (2\gamma\beta_K(S_K + Ks^2))$, but it vanishes as $K/T \rightarrow 1$. Beyond the boundary the only ways to hold all N assets are to reduce the covariance matrix softly (shrinkage toward a target, Ledoit and Wolf, 2004) or to impose a structure with fewer parameters (a factor model); both keep $K = N$ on the forecast side and pay a different, smaller price on the covariance side that Proposition 4 does not cover. The hard alternative is a basis of $K \ll T$ portfolios. The next subsection takes the comparison up.

2.5 Soft and hard on the covariance side

The covariance side has the same two reductions as the forecast side, and the same reason for them. This subsection first says where the covariance price comes from, then states the two reductions and what they cost, and closes with the size of the basis that the hard reduction should have.

Where the price comes from. Whiten the returns, so that $\Sigma = I$ and every eigenvalue of the true covariance matrix is one. The eigenvalues of the sample covariance matrix are not: with $N/T = y < 1$ they spread over the interval $[(1 - \sqrt{y})^2, (1 + \sqrt{y})^2]$ (Marčenko and Pastur, 1967), and for $y \geq 1$ a fraction $1 - 1/y$ of them is exactly zero. At $N = 100$, $T = 120$ the smallest sample eigenvalue is about $(1 - 0.91)^2 = 0.008$ times the truth. The position $\hat{\Sigma}^{-1}\hat{\mu}$ divides by these eigenvalues, so the directions whose sample variance came out smallest, by chance, receive weights that are two orders of magnitude too large. This is what the constants (18) measure: the average of $1/\hat{\lambda}$ over the spread is $1/(1 - y)$ and the average of $1/\hat{\lambda}^2$ is $1/(1 - y)^3$, which are α_K and β_K to first order. The price is paid in whichever directions come out with small sample variance, and after whitening these are random: this is why the constants multiply the whole of V^{-1} and scale every tradable direction alike. Nothing in this depends on the forecast.

Hard. Trade K basis portfolios and estimate only their $K \times K$ covariance matrix (Section 2.4). The number of estimated parameters falls from $N(N + 1)/2$ to $K(K + 1)/2$, the spread of the eigenvalues of $\hat{V}$ is governed by K/T instead of N/T , and the loss is (23): $\|a\|^2 - \varphi_K S_K + \beta_K K s^2$. The reduction is exact in the sense that Proposition 4

is an identity, and it has a hard limit: $K \leq T - 5$, and at unit length $\varphi_K > 0$, i.e. K/T below about 0.28. What it costs is the signal outside the basis, $\|a\|^2 - S_K$. The closest study of this trade-off is [Branger et al. \(2019\)](#). They sort assets by a criterion (expected excess returns, alphas, betas or the optimal weights), form equally weighted groups and hold the plug-in mean–variance portfolio of the groups; in a simplified setup they derive the optimal group size, which falls with the heterogeneity of the assets and rises with estimation risk, and they show that a sorting criterion that is itself estimated carries estimation risk of its own. In the terms of this section their strategy is a basis of $K = G$ group portfolios with the forecast used as it is, the rank-tiers cell of [Section 3](#), and their optimal group size is the counterpart of [Corollary 4](#). What the present framework adds is the separation of the basis from the reduction of the forecast inside it, the exact price of both ([Theorem 2](#)), and the link between the number of groups and the information coefficient ([Proposition 7](#)). Restricting the basis is itself a reduction of the covariance matrix. With G rank groups the position $D(D'\hat{\Sigma}D)^{-1}D'\hat{\mu}$ equals $\tilde{\Sigma}^{-1}\tilde{\mu}$ for any block-exchangeable $\tilde{\Sigma}$ (equal variances and equal covariances within each group, constant covariances between groups) with $D'\tilde{\Sigma}D = D'\hat{\Sigma}D$, and for $\tilde{\mu}$ the forecast averaged within groups: the position depends on $\hat{\Sigma}$ only through the $G(G+1)/2$ entries of the group covariance matrix. Related reductions are the constant-correlation target of [Ledoit and Wolf \(2004\)](#), the portfolio constraints of [Jagannathan and Ma \(2003\)](#) and the norm constraints of [DeMiguel et al. \(2009a\)](#), both of which act as covariance shrinkage, the hierarchical allocation of [Lopez de Prado \(2016\)](#), which imposes a block structure through clustering, and the correlation shrinkage of [Pedersen et al. \(2021\)](#). The difference is that the group covariance matrix is priced exactly in the investor’s loss ([Theorem 2](#) with $K = G$).

Soft. Keep all N assets and pull the sample covariance matrix toward a target with few parameters,

$$\tilde{\Sigma} = \delta F + (1 - \delta)\hat{\Sigma}, \quad \delta^* = \frac{\mathbb{E}\|\hat{\Sigma} - \Sigma\|^2}{\mathbb{E}\|\hat{\Sigma} - F\|^2}, \quad (25)$$

where F is a scaled identity, a constant-correlation matrix or a one-factor matrix and the norm is the Frobenius norm ([Ledoit and Wolf, 2004](#)); nonlinear shrinkage ([Ledoit and Wolf, 2017](#)) corrects each sample eigenvalue separately. The intensity has the form of κ^* in [Proposition 2](#), a variance of order N/T against variance plus squared bias, so that for N/T large $\tilde{\Sigma} \rightarrow F$ and the soft reduction collapses to its target, a hard reduction in disguise. In the language of [\(25\)](#) the hard reduction of the previous paragraph is shrinkage with intensity one toward a target whose $G(G+1)/2$ parameters are estimated; the soft reduction pulls with an intensity below one toward a fixed target. The hard reduction is priced exactly in the investor’s loss ([Theorem 2](#) with $K = G$); the soft one is priced in the Frobenius metric of Σ , not in the loss of [Lemma 1](#), and its optimality in that loss, together with shrinkage inside a basis, is the subject of a companion paper.

How many basis portfolios. Proposition 5 prices one addition at a time. A closed form for the size of the basis needs an assumption on how fast the tradable signal grows with K , and the natural one comes from a basis built on a ranking; it is stated at the end of Section 2.6, once the ranking is in place (Corollary 4).

2.6 The forecast within the action space

Where the argument stands. Section 2.3 showed that an investor who acts on fewer dimensions of his forecast can lose less, because every dimension he keeps costs s^2 and only the dimensions he drops cost signal: less can be more. Sections 2.4 and 2.5 showed the same for the covariance matrix: a basis of K portfolios is cheaper to estimate than N assets, and the price grows with K/T . The two reductions are different objects, one a choice of portfolios, the other a choice of how much of the forecast to believe, and an investor makes both. This section puts them together: given the basis, the forecast for the K basis portfolios can itself be used in part, and the question is how the two prices combine. The answer (Theorems 1 and 2) is that with a known covariance matrix the two prices add, the covariance side through the signal it makes untradable and the forecast side through the number of dimensions kept; with an estimated covariance matrix they interact, because β_K multiplies the noise of every forecast dimension kept.

Given the basis D , the investor still has a forecast $\hat{m} = D'\hat{\mu}$ for the K basis portfolios, and he may use less than all of it. Let Q be a $K \times K$ matrix applied to the forecast in return units, and

$$\tilde{Q} = V^{-1/2}QV^{1/2} \quad (26)$$

its image in risk units. The position is

$$w = Dx, \quad x = \frac{1}{\gamma}\hat{V}^{-1}Q\hat{m}, \quad (27)$$

i.e. $\gamma\tilde{v} = \tilde{Q}\tilde{b}$ when V is known. Q is the second choice and it is free: $Q = \kappa I$ is soft reduction inside the space, a projection is hard reduction, and $Q = I$ uses the forecast as it is. Section 2.3 is the case $D = I$.

Three reductions a ranking suggests. The reductions used in this paper, and by the familiar rules of Section 3, are the ones a forecast that ranks the basis portfolios suggests. Order the K basis portfolios by the forecast and let $z_1 \leq \dots \leq z_K$ be their standardized ranks. Three choices of Q keep successively less of the forecast's magnitudes while keeping its order:

- (i) *Steps.* Cut the ordering into G contiguous groups and replace each forecast by the average of its group: Q_G is group averaging, of rank G . The investor believes that

assets within a group have the same mean and that the G group means are what the data say they are.

- (ii) *Line*. Replace the G group averages by the fitted values of a regression of these averages on a constant and the group rank: $Q_{\text{line}} = X(X'X)^{-1}X'Q_G$ with $X = [1_K, \bar{z}_{g(\cdot)}]$ the $K \times 2$ matrix of a constant and each basis portfolio's group rank (or its weighted form, see Remark 7), of rank two; with equal groups the fitted values are those of the regression of the G group averages on the group rank. The investor believes that the mean rises linearly in the rank and keeps only a level and a slope; the magnitudes come from the data, the shape from the assumption.
- (iii) *Sign*. Keep only the direction top minus bottom with a magnitude fixed by convention: rank one, no magnitude from the data. This is the long-short sort, and the most reduced form of a ranking forecast; Guo and Wachter (2026) show for a single asset that even the sign of a forecast carries a premium. Between (ii) and (iii) sit the rank-weighted portfolios of Asness et al. (2013), the line at every rank without a covariance matrix.

The steps and the line are fixed matrices once the ranks are fixed, which is what Theorem 2 below requires, provided the ranking is not formed from the returns that enter $\hat{m}$; a ranking by the window's own sample means makes the partition a function of $\hat{m}$, and the condition of Section 3.8 applies. The sign is not a map of the forecast at all but a fixed position, the anchor x_0 of Corollary 2 with $Q = 0$. Together with $Q = I$ (every magnitude) and $Q = \kappa I$ (every magnitude, scaled) they are the “reduction” column of Table 1. The steps are treated in closed form below (Proposition 6 and Corollary 3); the line is the steps followed by one further projection, and the sort is the three-group steps with the middle group and the magnitudes removed.

Theorem 1 (Two-sided decomposition, known covariance). *For any D and any linear Q , with V known and $\tilde{b} = \tilde{a} + \tilde{\eta}$, $\text{Var}(\tilde{\eta}) = \tilde{\Psi}$,*

$$2\gamma \mathbb{E}[\text{loss}] = \underbrace{\|a\|^2 - \|\Pi_D a\|^2}_{\text{not tradable } (D)} + \underbrace{\|(I - \tilde{Q})\tilde{a}\|^2}_{\text{discarded inside } (Q)} + \underbrace{\text{tr}(\tilde{Q}\tilde{\Psi}\tilde{Q}')}_{\text{retained } (Q)}. \quad (28)$$

If $\tilde{Q}$ is an orthogonal projection of rank k and $\tilde{\Psi} = s^2 I$, the last term is ks^2 .

Proof. Start from Lemma 2: the loss is the untradable term plus $\|\gamma\tilde{v} - \tilde{a}\|^2$. With V known, $x = V^{-1}Q\hat{m}/\gamma$, so $\gamma\tilde{v} = \gamma V^{1/2}x = V^{-1/2}Q\hat{m} = V^{-1/2}QV^{1/2}V^{-1/2}\hat{m} = \tilde{Q}\tilde{b}$. Insert $\tilde{b} = \tilde{a} + \tilde{\eta}$:

$$\gamma\tilde{v} - \tilde{a} = \tilde{Q}\tilde{a} + \tilde{Q}\tilde{\eta} - \tilde{a} = -(I - \tilde{Q})\tilde{a} + \tilde{Q}\tilde{\eta}.$$

Expand the squared norm:

$$\|(I - \tilde{Q})\tilde{a}\|^2 - 2\tilde{a}'(I - \tilde{Q})'\tilde{Q}\tilde{\eta} + \|\tilde{Q}\tilde{\eta}\|^2.$$

The middle term is linear in $\tilde{\eta}$ and $\mathbb{E}[\tilde{\eta}] = 0$, so it vanishes in expectation (for a general linear $\tilde{Q}$ it does not vanish realization by realization, unlike in Proposition 3; this is why the theorem is stated for the expected loss). The last term is $\mathbb{E}[\tilde{\eta}'\tilde{Q}'\tilde{Q}\tilde{\eta}] = \text{tr}(\tilde{Q}'\tilde{Q}\tilde{\Psi}) = \text{tr}(\tilde{Q}\tilde{\Psi}\tilde{Q}')$ by the trace trick. If $\tilde{\Psi} = s^2 I$ this is $s^2 \text{tr}(\tilde{Q}'\tilde{Q}) = s^2 \|\tilde{Q}\|_F^2$, the squared Frobenius norm. For an orthogonal projection $\tilde{Q}'\tilde{Q} = \tilde{Q}$ and $\text{tr} \tilde{Q} = \text{rank} \tilde{Q} = k$. Finally, the untradable term is $\|a\|^2 - \|\Pi_D a\|^2$ from Lemma 2. $\square$

With V known the two reductions are charged separately and do not interact: the basis decides what is tradable, the map Q decides how much of the tradable signal is used and how many dimensions of noise come with it. The noise count is the rank of $\tilde{Q}$, which is at most K ; the price of each dimension is still s^2 .

Remark 7 (Oblique reductions). Q is specified in return units (an average, a fitted line, a scalar), and its risk-unit image $\tilde{Q}$ is an orthogonal projection only if Q commutes with V . Otherwise $\tilde{Q}$ is an oblique projection and $\|\tilde{Q}\|_F^2 > \text{rank} \tilde{Q}$: an idempotent matrix has nonzero singular values of at least one, with equality throughout only when it is orthogonal. A reduction that is not aligned with the risk metric therefore keeps more than s^2 per dimension. For averaging over groups of basis portfolios the condition is that V maps each group indicator into the span of the group indicators, $V1_g \in \text{span}\{1_h\}$ for every g ; this holds when V is block-exchangeable (equal variances and equal correlations within a group, constant covariances between any two groups), and then $\text{span}\{1_g\}$ is V -invariant and the group-mean reduction is orthogonal in risk units. The same condition, applied to Σ and the asset groups, is what makes the restriction to equal-weighted group portfolios ($D = \text{group indicators}$) coincide with the group-mean projection in risk units, $\text{span}(\Sigma^{1/2}D) = \text{span}\{1_g\}$ (Appendix A.1).

Theorem 2 (Two-sided decomposition, estimated covariance). *Under Assumption 1, for a fixed Q whose risk-unit image $\tilde{Q}$ is an orthogonal projection of rank k , the position (27) has*

$$2\gamma \mathbb{E}[\text{loss}] = (\|a\|^2 - \|\Pi_D a\|^2) + (\|\tilde{a}\|^2 - \|\tilde{Q}\tilde{a}\|^2) + (1 - 2\alpha_K + \beta_K) \|\tilde{Q}\tilde{a}\|^2 + \beta_K k s^2. \quad (29)$$

For general linear Q the last three terms read $\|\tilde{a}\|^2 - 2\alpha_K \tilde{a}'\tilde{Q}\tilde{a} + \beta_K \|\tilde{Q}\tilde{a}\|^2 + \beta_K s^2 \|\tilde{Q}\|_F^2$.

Proof. The proof follows Proposition 4 with $Q\hat{m}$ in place of $\hat{m}$. Now $x = \hat{V}^{-1}Q\hat{m}/\gamma$ and $\text{CE}(Dx) = x'm - \frac{\gamma}{2}x'Vx$.

Mean term. $x'm = \hat{m}'Q'\hat{V}^{-1}m/\gamma$. Expectation over $\hat{V}$ first (independence), then over $\hat{m}$: $\mathbb{E}[x'm] = \alpha_K m'Q'V^{-1}m/\gamma$. To express this in risk units insert $m = V^{1/2}\tilde{a}$ and $V^{-1}m = V^{-1/2}\tilde{a}$:

$$m'Q'V^{-1}m = \tilde{a}'V^{1/2}Q'V^{-1/2}\tilde{a} = \tilde{a}'\tilde{Q}'\tilde{a} = \tilde{a}'\tilde{Q}\tilde{a},$$

since $V^{1/2}Q'V^{-1/2} = (V^{-1/2}QV^{1/2})' = \tilde{Q}'$ and a scalar equals its transpose.

Variance term. $x'Vx = \hat{m}'Q'\hat{V}^{-1}V\hat{V}^{-1}Q\hat{m}/\gamma^2$. Over $\hat{V}$ first:

$$\mathbb{E}[x'Vx] = \beta_K \mathbb{E}[\hat{m}'Q'V^{-1}Q\hat{m}]/\gamma^2.$$

Split $\hat{m} = m + (\hat{m} - m)$ as before:

$$\mathbb{E}[\hat{m}'Q'V^{-1}Q\hat{m}] = m'Q'V^{-1}Qm + \text{tr}(Q'V^{-1}QV/T).$$

The first term is $\|\tilde{Q}\tilde{a}\|^2$ by the same substitution as above. For the second, using the cyclic property of the trace,

$$\text{tr}(Q'V^{-1}QV) = \text{tr}(V^{1/2}Q'V^{-1/2}V^{-1/2}QV^{1/2}) = \text{tr}(\tilde{Q}'\tilde{Q}) = \|\tilde{Q}\|_F^2.$$

Collect. $2\gamma \mathbb{E}[\text{CE}(Dx)] = 2\alpha_K \tilde{a}'\tilde{Q}\tilde{a} - \beta_K \|\tilde{Q}\tilde{a}\|^2 - \beta_K s^2 \|\tilde{Q}\|_F^2$. Subtract from $\|a\|^2$ and add and subtract $\|\tilde{a}\|^2 = \|\Pi_D a\|^2$:

$$2\gamma \mathbb{E}[\text{loss}] = (\|a\|^2 - \|\tilde{a}\|^2) + \|\tilde{a}\|^2 - 2\alpha_K \tilde{a}'\tilde{Q}\tilde{a} + \beta_K \|\tilde{Q}\tilde{a}\|^2 + \beta_K s^2 \|\tilde{Q}\|_F^2,$$

which is the general form. If $\tilde{Q}$ is an orthogonal projection, $\tilde{a}'\tilde{Q}\tilde{a} = \tilde{a}'\tilde{Q}'\tilde{Q}\tilde{a} = \|\tilde{Q}\tilde{a}\|^2$ and $\|\tilde{Q}\|_F^2 = k$, so the middle three terms become $\|\tilde{a}\|^2 - 2\alpha_K \|\tilde{Q}\tilde{a}\|^2 + \beta_K \|\tilde{Q}\tilde{a}\|^2$; adding and subtracting $\|\tilde{Q}\tilde{a}\|^2$ turns this into $(\|\tilde{a}\|^2 - \|\tilde{Q}\tilde{a}\|^2) + (1 - 2\alpha_K + \beta_K)\|\tilde{Q}\tilde{a}\|^2$. $\square$

Theorem 2 is the general statement of this section. The Wishart constants carry the dimension of the covariance matrix, K ; the noise count carries the dimension of the forecast, $k \leq K$. The two sides can be chosen separately, but with an estimated covariance matrix their prices interact: β_K scales the noise of every retained forecast dimension, and the inversion error scales the signal retained. A small K keeps α_K, β_K near one, a small k keeps the noise count low, and the cost of each is the signal it makes unreachable. The theorem contains as special cases the plug-in ($D = I, Q = I$), the soft rule inside any basis ($Q = \kappa I$), any hard rule inside any basis, and the two-sided reductions of Section 3. With $D = I$ and $Q = \kappa I$ the theorem is the expected-loss formula of Kan and Zhou (2007); the extension is to an arbitrary basis and an arbitrary fixed reduction inside it.

Both theorems are confirmed by Monte Carlo in the supplementary code; for oblique $\tilde{Q}$ the general form holds and the projection form with the count k does not.

Two of the familiar rules are not of the form (27): they add a fixed position that does not depend on the forecast. The theorem extends to them.

Corollary 2 (Affine positions). *Let $x = x_0 + \gamma^{-1}\hat{V}^{-1}Q\hat{m}$ with a fixed $x_0 \in \mathbb{R}^K$, and let $u = \gamma V^{1/2}x_0$ be the anchor in the risk units of the basis. Under Assumption 1,*

$$2\gamma \mathbb{E}[\text{loss}] = (\|a\|^2 - \|\tilde{a}\|^2) + \|\tilde{a} - u\|^2 - 2\alpha_K (\tilde{a} - u)' \tilde{Q} \tilde{a} + \beta_K \|\tilde{Q} \tilde{a}\|^2 + \beta_K s^2 \|\tilde{Q}\|_F^2. \quad (30)$$

With V known ($\alpha_K = \beta_K = 1$) this is $(\|a\|^2 - \|\tilde{a}\|^2) + \|(I - \tilde{Q})\tilde{a} - u\|^2 + s^2 \|\tilde{Q}\|_F^2$: the anchor absorbs part of the signal that the reduction discards and adds no noise.

Proof. As in Theorem 2, with three additional terms from x_0 . In the mean term, $\mathbb{E}[x'_0 m] = x'_0 m = u' \tilde{a} / \gamma$ because $m = V^{1/2} \tilde{a}$ and $u = \gamma V^{1/2} x_0$. In the variance term, $x' V x$ gains $x'_0 V x_0 = \|u\|^2 / \gamma^2$ and the cross term $2x'_0 V \mathbb{E}[\hat{V}^{-1} Q \hat{m}] / \gamma = 2\alpha_K x'_0 Q m / \gamma = 2\alpha_K u' \tilde{Q} \tilde{a} / \gamma^2$, using $\mathbb{E}[\hat{V}^{-1}] = \alpha_K V^{-1}$ and $x'_0 V V^{-1} Q m = x'_0 Q V^{1/2} \tilde{a} = \gamma^{-1} u' V^{-1/2} Q V^{1/2} \tilde{a}$. Collecting, $2\gamma \mathbb{E}[\text{CE}(Dx)]$ gains $2u' \tilde{a} - \|u\|^2 - 2\alpha_K u' \tilde{Q} \tilde{a}$, and $\|\tilde{a}\|^2 - 2u' \tilde{a} + \|u\|^2 = \|\tilde{a} - u\|^2$. For known V , $\|\tilde{a} - u\|^2 - 2(\tilde{a} - u)' \tilde{Q} \tilde{a} + \|\tilde{Q} \tilde{a}\|^2 = \|\tilde{a} - u - \tilde{Q} \tilde{a}\|^2$. $\square$

The anchor is a position the investor holds regardless of the forecast: the equal-weight portfolio in the three-fund rule of [Tu and Zhou \(2011\)](#) and in the parametric policy of [Brandt et al. \(2009\)](#) (Section 3). It can lower the loss only through the signal term, by covering part of what $\tilde{Q}$ discards; it has no noise of its own.

Remark 8 (Data-dependent reductions). Theorem 2 takes Q as a fixed matrix, chosen before the window's data are seen. The proof uses this once, when $\mathbb{E}[Q\hat{m}] = Qm$ is inserted; if Q depends on $\hat{m}$ or $\hat{V}$ this step fails. The three reductions of the paragraph above, steps, line and sign, are fixed: their matrices depend on the ranks only, not on the values, provided the ranking is not formed from the returns that enter $\hat{m}$; a ranking by the window's own sample means makes the partition a function of $\hat{m}$, and the condition of Section 3.8 applies. A scalar estimated from the window, or a number of groups chosen from a correlation measured in the window, is not. Such a choice is a fourth estimated quantity, and its cost is not in (29). For the scalar the cost is known ([Kan and Zhou, 2007](#)) and is small when k and K are small; Section 2.7 uses it. For a number of groups chosen from the same returns that enter $\hat{m}$, the choice and the forecast error are dependent, and the usual repair is to measure the choice on a separate window or by cross-fitting; Section 4 does so. The same applies to the basis: Assumption 1 requires D to be fixed, and a basis rebuilt every month from a ranking is a sequence of bases, each fixed at its own date, whose histories are returns of managed portfolios; Section 4 states the assumption

under which the theorems apply to them. The theorems of this section price every cell of the grid; what they do not price is the selection of the cell from the data.

Rank groups. One basis and one reduction deserve a closed form because they are the ones a ranking forecast suggests. Suppose the forecast orders the K basis portfolios (with $D = I$, the assets), and suppose for this paragraph that they are uncorrelated, V diagonal, so that coordinates in risk units correspond one to one to basis portfolios (correlated bases: Remark 7 and Appendix A.1). Let $\mathcal{G}$ be a partition of them into G contiguous groups, and let $\tilde{Q}_{\mathcal{G}}$ be the orthogonal projection, in risk units, onto the span of the group indicators, so that $(\tilde{Q}_{\mathcal{G}}\tilde{a})_i = \bar{a}_{g(i)}$.

Proposition 6 (Loss of a partition). *With $\tilde{\Psi} = s^2 I$, the term $\|(I - \tilde{Q}_{\mathcal{G}})\tilde{a}\|^2 + ks^2$ in Theorem 1 equals*

$$\sum_{g=1}^G \sum_{i \in g} (\tilde{a}_i - \bar{a}_g)^2 + G s^2. \quad (31)$$

The noise depends on $\mathcal{G}$ only through G ; the signal discarded is the within-group dispersion of the profile $\tilde{a}$, minimized for given G by the partition with the smallest within-group sum of squares (one-dimensional k -means).

Proof. $\tilde{Q}_{\mathcal{G}}$ is an orthogonal projection onto the span of the G group indicators, which are orthogonal to each other (no two groups overlap), so its rank is G and the noise term in Theorem 1 is Gs^2 . For the signal term, $(I - \tilde{Q}_{\mathcal{G}})\tilde{a}$ has i -th entry $\tilde{a}_i - \bar{a}_{g(i)}$, the deviation of element i from the mean of its group; its squared norm is the sum of these squared deviations over all elements, grouped by g . For a fixed number of groups the noise term is constant, so minimizing the loss over partitions means minimizing the within-group sum of squares, which is the k -means objective in one dimension. $\square$

Groups may be of any size and cut anywhere. If the profile is flat inside and differs from its level only at the two ends, the best three-group partition is two small tail groups and one large middle group. If the profile is linear in the rank, equal groups are optimal and G has a closed form.

Corollary 3 (Linear profile, equal groups). *Let $\tilde{\Psi} = s^2 I$ (isotropic error, Remark 1), let $\tilde{a}_i = \bar{a} + \tau z_i$ with z_i the standardized rank (equally spaced, mean zero, unit variance) and $\mathcal{G}$ consist of G equal contiguous groups. Then the within-group sum of squares is $K\tau^2 \frac{(K/G)^2 - 1}{K^2 - 1} \approx K\tau^2/G^2$, so*

$$2\gamma \mathbb{E}[\text{loss}](G) \approx \frac{K\tau^2}{G^2} + G s^2, \quad G^* = \left(\frac{2K\tau^2}{s^2} \right)^{1/3} = (2K\tau^2 T)^{1/3} \quad (32)$$

up to the untradable term, which does not depend on G .

Proof. The within-group sum of squares of $\tilde{a}$ equals τ^2 times the within-group sum of squares of z , because $\tilde{a}_i - \bar{\tilde{a}}_g = \tau(z_i - \bar{z}_g)$. So it suffices to compute the latter. Let the z_i be equally spaced with spacing d . A standard identity says that n equally spaced points with spacing d have a sum of squared deviations from their mean of $d^2 n(n^2 - 1)/12$. Applied to all K points, the unit-variance normalization $\sum_i (z_i - \bar{z})^2 = K$ gives $d^2 K(K^2 - 1)/12 = K$, i.e. $d^2 = 12/(K^2 - 1)$. Applied to one group of $m = K/G$ consecutive points, the within-group sum of squares is $d^2 m(m^2 - 1)/12$; summing over the G groups gives $d^2 Gm(m^2 - 1)/12 = d^2 K(m^2 - 1)/12$. Substituting d^2 :

$$\sum_g \sum_{i \in g} (z_i - \bar{z}_g)^2 = K \frac{m^2 - 1}{K^2 - 1} = K \frac{(K/G)^2 - 1}{K^2 - 1} \approx \frac{K}{G^2}$$

for K large. Multiply by τ^2 and add the noise term Gs^2 from Proposition 6. The function $K\tau^2/G^2 + Gs^2$ is convex in $G > 0$; its derivative $-2K\tau^2/G^3 + s^2$ vanishes at $G^3 = 2K\tau^2/s^2$. $\square$

The optimum balances a signal loss falling as G^{-2} against a noise cost rising linearly in G , and grows with the cube root of the cross-section, of the slope of the profile and of the window: over the empirical range it is a small number.

The cube root is the familiar rate of a partitioning estimator, whose squared bias falls as G^{-2} while its variance rises with G . Cattaneo et al. (2020) derive it for characteristic-sorted portfolios, with an optimal number of portfolios proportional to $(nT)^{1/3}$ for n assets and T periods, and find optimal numbers between ten and several hundred. Corollary 3 differs in the loss and therefore in the price. Their criterion is the mean squared error of the estimated return–characteristic relation, in which the variance of a portfolio mean falls with the number of assets in it. Here the criterion is the investor’s certainty equivalent, in which every retained dimension costs $s^2 = 1/T$ of squared Sharpe ratio regardless of how many assets it averages: the estimated Sharpe ratio of any fixed portfolio has variance $1/T$, because averaging lowers the noise of the mean and the risk of the portfolio in the same proportion. The price per group therefore does not fall with n/G , and G^* is small.

One extension is stated without proof here: the line of reduction (ii), which replaces the G group means by the fitted values of a regression on a constant and the rank, is a further projection of rank two inside the group space, with noise $2s^2$ instead of Gs^2 and a signal term that is the deviation of the profile from a line. The $2s^2$ holds for the line fitted in the metric of V (generalized least squares), which is orthogonal in risk units; a line fitted by ordinary least squares in return units is a fixed Q but oblique in risk units, and keeps slightly more than $2s^2$ (Remark 7; the difference is small in the cases we compute). The

other extension, the covariance matrix estimated in the group space so that $K = G$, is Corollary 4 below.

From the slope to the information coefficient. The investor does not observe τ . What he can measure is how well the forecast ranks: the cross-sectional correlation between the forecast's rank and the return of the following month, averaged over the window, which the practitioner literature calls the information coefficient (Grinold, 1989). The next proposition says that this correlation determines τ , so that every input of (32) is observable.

Proposition 7 (Slope and information coefficient). *Let the setting of Corollary 3 hold (basis uncorrelated, $\tilde{a}_i = \bar{a} + \tau z_i$ with z_i the standardized rank of the forecast), and let the realized return of element i over one month, in risk units, be $\tilde{r}_i = \tilde{a}_i + \varepsilon_i$ with ε_i of mean zero and unit variance, uncorrelated with z_i across i . Then the cross-sectional correlation between z and $\tilde{r}$ is*

$$\rho = \frac{\tau}{\sqrt{1 + \tau^2}}, \quad \text{i.e.} \quad \tau^2 = \frac{\rho^2}{1 - \rho^2}, \quad (33)$$

and Corollary 3 becomes

$$G^* = \left(\frac{2K\tau^2}{s^2} \right)^{1/3} = \left(\frac{2KT\rho^2}{1 - \rho^2} \right)^{1/3}. \quad (34)$$

Proof. Take cross-sectional moments over the K elements, with z standardized so that its mean is zero and its variance one. The covariance between z and $\tilde{r}$ is $\text{cov}(z, \tilde{r}) = \text{cov}(z, \bar{a} + \tau z) + \text{cov}(z, \varepsilon) = \tau \text{var}(z) + 0 = \tau$. The variance of $\tilde{r}$ is $\text{var}(\tilde{a}) + \text{var}(\varepsilon) + 2 \text{cov}(\tilde{a}, \varepsilon) = \tau^2 + 1 + 0$, since $\text{var}(\tilde{a}) = \tau^2 \text{var}(z) = \tau^2$ and the noise has unit variance in risk units by construction of the units. Hence $\rho = \tau / \sqrt{\tau^2 + 1}$. Squaring and solving, $\rho^2(1 + \tau^2) = \tau^2$ gives $\tau^2 = \rho^2 / (1 - \rho^2)$. Insert into (32) with $s^2 = 1/T$. $\square$

Four remarks on what (33) assumes and what the investor measures. First, the correlation is between the forecast's rank and the return *in risk units*, i.e. the return of each basis element divided by its volatility; with raw returns and unequal volatilities the correspondence is approximate, and the application measures the coefficient on volatility-scaled returns. Second, the information coefficient of practice is a rank correlation (Spearman), which ranks the returns as well; for the small values that return forecasts attain the two coincide to within a few percent, since for jointly normal variables the Spearman correlation is $(6/\pi) \arcsin(\rho/2) \approx 0.95\rho$. Third, one month's cross-sectional correlation over K elements has a standard error of about $1/\sqrt{K}$, so the trailing average over T months estimates ρ with a standard error of about $1/\sqrt{KT}$, below 0.01 for a hundred assets and ten years; the estimate is nonetheless a quantity measured in the window, and a G set from it is a data-dependent reduction in the sense of Remark 8. The earlier working form,

which replaced τ^2 by ρ^2 and s^2 by $(1 - \rho^2)/T$, gives the same ratio τ^2/s^2 and hence the same G^* ; the proposition says which of the two quantities carries the $1 - \rho^2$. Fourth, the proposition assumes that the forecast ranks the true Sharpe ratios without error, so that $\rho < 1$ only because monthly returns are noisy. A ranking with its own error, $\tilde{a}_i = \tilde{a} + \tau z_i + \nu_i$ with ν uncorrelated with z and of cross-sectional variance σ_ν^2 , gives $\rho = \tau/\sqrt{1 + \tau^2 + \sigma_\nu^2}$, i.e. $\tau^2 = \rho^2(1 + \sigma_\nu^2)/(1 - \rho^2)$. Since σ_ν^2 is a dispersion of monthly squared Sharpe ratios, the factor $1 + \sigma_\nu^2$ is negligible and the measured ρ still determines τ . The part ν of the signal that the ranking misses is not entirely lost to the grouping: if it is unrelated to the ranks, a partition into G groups retains $G\sigma_\nu^2$ of it in expectation, the effective price per group falls from s^2 to $s^2 - \sigma_\nu^2$, and $G^* = (2K\tau^2/(s^2 - \sigma_\nu^2))^{1/3}$. This exceeds (34) noticeably only when σ_ν^2 is not small relative to $1/T$; when $\sigma_\nu^2 \geq s^2$ every additional group pays, and a partition of the ranking is no longer the right reduction. (Monte Carlo check: supplementary code, part 5.)

Both sides at once. Corollary 3 took the basis as given and chose the number of groups on the forecast side alone. When the group portfolios are also the basis in which the covariance matrix is estimated, D has $K = G$ columns, and the within-group dispersion $K\tau^2/G^2$ of Corollary 3 is no longer a term discarded by Q but the untradable term of Proposition 4: with G groups the tradable signal is $S_G = S_\infty - K\tau^2/G^2$, where K is the number of ranked elements and S_∞ the signal tradable with every one of them in the basis. The choice of G then carries the covariance price as well, and Section 2.5 left the closed form for this case.

Corollary 4 (Size of the basis). *Let $D_1 \subset D_2 \subset \dots$ be a nested sequence of bases whose tradable signal satisfies $S_K = S_\infty - C/K^2$ for constants $S_\infty, C > 0$, let the forecast inside the basis be used as it is ($Q = I$), and consider the position at unit length. To first order in K/T , the loss (23) is*

$$2\gamma \mathbb{E}[\text{loss}](K) \approx \|a\|^2 - S_\infty + \frac{C}{K^2} + (1 + S_\infty) \frac{K}{T}, \quad (35)$$

minimized at

$$K^* = \left(\frac{2CT}{1 + S_\infty} \right)^{1/3}. \quad (36)$$

Proof. Expand the constants for K/T small: $\alpha_K = (T - 1)/(T - K - 2) \approx 1 + K/T$ and $\beta_K \approx (1 - K/T)^{-3} \approx 1 + 3K/T$, so $\varphi_K = 2\alpha_K - \beta_K \approx 1 - K/T$ and $\beta_K K s^2 \approx (K/T)(1 + 3K/T)$. Insert into (23) with $S_K = S_\infty - C/K^2$:

$$\|a\|^2 - \left(1 - \frac{K}{T}\right) \left(S_\infty - \frac{C}{K^2}\right) + \frac{K}{T} + \frac{3K^2}{T^2} = \|a\|^2 - S_\infty + \frac{C}{K^2} + \frac{K}{T}(1 + S_\infty) - \frac{C}{KT} + \frac{3K^2}{T^2}.$$

The last two terms are of smaller order than the two before them (their ratios are K/T and $3K/T$) and are dropped. The remaining function of K is convex with derivative $-2C/K^3 + (1 + S_\infty)/T$, which vanishes at $K^3 = 2CT/(1 + S_\infty)$. $\square$

For rank groups with a linear profile, $C = K\tau^2$ (with K now the number of ranked elements, fixed) and Corollary 4 is Corollary 3 with the price of a dimension raised from s^2 to $s^2(1 + S_\infty)$: the extra $S_\infty s^2$ is the inversion error, the first-order effect of $\varphi_K < 1$ on the signal already tradable, and it is small when the attainable squared Sharpe ratio is small. So when the forecast is used as it is, estimating the covariance matrix in the group space changes the optimal number of groups little. The picture changes when the forecast inside the basis is reduced further, for instance to the line mentioned above, so that the noise count stops growing with the number of groups: the loss is then $C/G^2 + S_\infty G/T$ up to constants, the covariance price alone bounds the basis, and $G^* = (2CT/S_\infty)^{1/3}$, which for $C \approx S_\infty$ (signal mostly in the ranking) is $(2T)^{1/3}$, about six at $T = 120$. In that case the number of groups is set by the covariance side, not by the forecast side. With the optimal scalar of Section 2.7 in place of unit length the first-order form is the same, with S_∞ replaced by $S_\infty^2/(S_\infty + ks^2)$: the scalar removes the leverage part $(\alpha_K - 1)^2$ of the inversion error, which is of second order, but not the direction part $\beta_K - \alpha_K^2 \approx (K + 1)/T$.

2.7 Direction and length

Every position has a direction and a length. The choices of D and Q fix the direction; the length is a separate problem, and the only one on which the certainty equivalent and the Sharpe ratio disagree.

Proposition 8 (Angle identity). *Let $\gamma v = \kappa d$ with $\|d\| = 1$, $\kappa \geq 0$. Then $2\gamma \text{loss} = \|a\|^2 - 2\kappa d'a + \kappa^2$, minimized at $\kappa^\circ = d'a$ with*

$$2\gamma \text{loss}(\kappa^\circ) = \|a\|^2(1 - \cos^2(d, a)), \quad \text{Sharpe ratio of } d = d'a = \|a\| \cos(d, a). \quad (37)$$

Proof. Lemma 1 with $\gamma v = \kappa d$: $2\gamma \text{loss} = \|\kappa d - a\|^2 = \kappa^2\|d\|^2 - 2\kappa d'a + \|a\|^2$, and $\|d\| = 1$. This is a convex quadratic in κ with minimum at $\kappa^\circ = d'a$. Substituting, $\|a\|^2 - 2(d'a)^2 + (d'a)^2 = \|a\|^2 - (d'a)^2$. By the definition of the angle between two vectors, $d'a = \|d\|\|a\|\cos(d, a) = \|a\|\cos(d, a)$, so $\|a\|^2 - (d'a)^2 = \|a\|^2(1 - \cos^2(d, a))$. For the Sharpe ratio: the position w with $\Sigma^{1/2}w \propto d$ has expected excess return $w'\mu = (\Sigma^{1/2}w)'a$ and standard deviation $\|\Sigma^{1/2}w\|$, so its Sharpe ratio is $d'a$. $\square$

At the optimal length the loss depends on the direction only through its angle with the truth, and the Sharpe ratio of the position is the maximal Sharpe ratio times the cosine

of that angle. A Sharpe-ratio investor solves the direction problem only; a certainty-equivalent investor solves both. The Sharpe-ratio gap $\theta - d'a = \theta(1 - \cos(d, a))$ is the direction-only counterpart of the loss identity, and the bound of Salcher et al. (2026), Sharpe-ratio gap $\leq c_0 \cdot \text{RAFE} + \theta \cdot \text{C-RAFE}$, is obtained from it by normalizing $\hat{C}\hat{b}$ to unit length and applying the triangle inequality to $\|d - a/\theta\|$: the normalization removes the length, which the Sharpe ratio does not see, and the triangle inequality separates the forecast error from the covariance error (Remark 5). The certainty equivalent keeps the length, which is why Lemma 1 is an identity where the Sharpe-ratio statement is a bound.

The investor does not know $d'a$. For a position with k retained dimensions he observes $\|\tilde{Q}\tilde{b}\|^2$, whose expectation is $\|\tilde{Q}\tilde{a}\|^2 + ks^2$ by Theorem 1: the estimated squared Sharpe ratio overstates the true one by exactly the noise retained. Subtracting it gives the feasible form of the scalar in Remark 2,

$$\hat{\kappa} = \frac{(\|\tilde{Q}\tilde{b}\|^2 - ks^2)^+}{\|\tilde{Q}\tilde{b}\|^2}, \quad (38)$$

which for $D = I$, $Q = I$ is the two-fund scalar of Kan and Zhou (2007) with known Σ . Equation (38) presumes V known and $\tilde{Q}$ orthogonal. With $\hat{V}$ the optimal scalar is $(\alpha_K/\beta_K) \|\tilde{Q}\tilde{a}\|^2/(\|\tilde{Q}\tilde{a}\|^2 + ks^2)$ (from the general form of Theorem 2 with $\tilde{Q}$ replaced by $\kappa\tilde{Q}$), which contains the unknown $\|\tilde{Q}\tilde{a}\|^2$ and is therefore an oracle value; the feasible version replaces it by a bias-corrected estimate, as in Kan and Zhou (2007), and the explicit form is in Appendix A.2. Two properties matter. The scalar corrects the length, not the direction. And a position with small k and K has less to correct, because ks^2 and α_K, β_K all fall.

2.8 Summary

An investor who acts on a forecast makes three choices.

1. *The action space D* : which K portfolios to trade. It fixes the dimension in which the covariance matrix is estimated and the signal that is tradable at all (Proposition 4). Its price is α_K, β_K , a function of K and T only.
2. *The reduction of the forecast inside the space Q* : how many and which of the K dimensions of the forecast to act on, hard or soft (Theorem 1). Its price is s^2 per dimension retained, amplified by β_K when the covariance matrix is estimated (Theorem 2).
3. *The length* (Proposition 8): a scalar that the Sharpe ratio does not see and the certainty equivalent does.

Table 1: Familiar portfolios as cells of the grid. D : the basis portfolios, K their number, the dimension in which the covariance matrix is estimated. Q : the reduction of the forecast inside the basis, k its rank; “soft” means a scalar on all K dimensions. Length: fixed (unit or volatility target) or a data-driven scalar. The last column says whether K or k change with the forecast’s measured quality.

Rule	Source	Forecast	Basis D (K)	Reduction Q (k)	Length	Moves?
Plug-in	Markowitz (1952); Jobson and Korkie (1980)	sample mean	assets (N)	I (N)	unit	no
Kan-Zhou two-fund	Kan and Zhou (2007)	sample mean	assets (N)	soft, κI	κ^*	no
Kan-Zhou three-fund	Kan and Zhou (2007)	sample mean	assets (N)	soft, two scalars (level, cross-section)	two scalars	no
Tu-Zhou	Tu and Zhou (2011)	sample mean	assets (N)	three-fund plus fixed $1/N$ (affine)	three scalars	no
Minimum variance	Jagannathan and Ma (2003); DeMiguel et al. (2009a)	none	assets (N)	level only (1)	fixed	no
$1/N$	DeMiguel et al. (2009b)	none	equal-weight portfolio (1)	level (1)	fixed	no
Decile sort	Fama and French (1993); Gu et al. (2020)	ranking	top minus bottom decile (1)	sign (1)	fixed	no
Parametric tilt (BSV)	Brandt et al. (2009)	characteristic z	$1/N$ and z -tilt (2)	slope only (1)	one scalar	no
Rank tiers	Branger et al. (2019); Cattaneo et al. (2020)	ranking	G group portfolios (G)	I (G)	scalar	no (G fixed)
Factor covariance	Chan et al. (1999); Fama and French (1993)	any	assets, structured Σ	any	any	no
Asset clusters	Lopez de Prado (2016)	sample mean	K industry or cluster portfolios (K)	I (K)	scalar	no (K fixed)
Ledoit-Wolf	Ledoit and Wolf (2004); Ledoit and Wolf (2017)	any	assets, soft Σ	any	any	no
Principal components	Kozak et al. (2020); Lettau and Pelger (2020)	any	top- K eigenvectors (K)	any	any	no
Section 4	this paper	any	G group portfolios	line (2)	$\kappa^*(G, T)$	G from ρ

The two reductions nest and are priced by different quantities, K in the Wishart constants and k in the noise count, which interact only through β_K multiplying the retained noise. Both are hard choices (which dimensions) with soft alternatives (how much of every dimension), and the forecast itself, historical mean or otherwise, is an input to all three. The rules of the portfolio-choice literature are fixed settings of (D, Q, length) ; the next section makes the correspondence explicit.

3 Familiar portfolios as cells

This section places the portfolio rules used in the literature and in practice on the grid of Section 2.8. For each rule we state the forecast it takes as input, the cell it occupies (the basis D with its dimension K , the reduction Q with its rank k , and the length), and its expected loss as given by the theorems above. The purpose is not to rank the rules, which depends on the forecast and the data, but to show that they are points of one grid and that every one of them is a fixed point: each sets K and k once, independently of what the forecast can deliver. Table 1 summarizes.

3.1 The plug-in portfolio

Definition. $w = \hat{\Sigma}^{-1}\hat{\mu}/\gamma$ with the sample moments.

Cell. $D = I$ ($K = N$), $Q = I$ ($k = N$), forecast the sample mean, length one.

Loss. Theorem 2 with nothing discarded: $(1 - 2\alpha_N + \beta_N)\theta^2 + \beta_N N s^2$. Both untradable and discarded terms are zero; the inversion error and the amplified noise are paid in full, at N dimensions each. It is the origin of the grid, the cell from which every other rule departs in one or more coordinates; it does not exist for $N \geq T$ and has infinite expected loss for $N \geq T - 4$ (Remark 6).

3.2 The Kan–Zhou two-fund rule

Definition. $w = \kappa \hat{\Sigma}^{-1} \hat{\mu} / \gamma$, κ chosen to maximize expected out-of-sample utility (Kan and Zhou, 2007).

Cell. $D = I$, $Q = \kappa I$: soft reduction on all N dimensions inside the full basis; length κ^* .

Loss. The general form of Theorem 2 with $\tilde{Q} = \kappa I$: $\theta^2 - 2\alpha_N \kappa \theta^2 + \beta_N \kappa^2 (\theta^2 + N s^2)$, minimized at $\kappa^* = (\alpha_N / \beta_N) \theta^2 / (\theta^2 + N/T)$, the Kan–Zhou scalar up to the normalization of $\hat{\Sigma}$. The rule keeps every dimension of forecast and covariance matrix and corrects the length: it solves the problem of Section 2.7 for the fixed direction “all N ” and leaves the direction where the plug-in has it. The scalar depends on the forecast only through θ and N/T because the forecast is the sample mean, whose error is pinned at I/T .

3.3 The Kan–Zhou three-fund rule and Tu–Zhou

Definition. Kan–Zhou three-fund: $w = (c/\gamma) [\eta \hat{\Sigma}^{-1} \hat{\mu} + (1 - \eta) \hat{\mu}_g \hat{\Sigma}^{-1} 1]$ with $\hat{\mu}_g = 1' \hat{\Sigma}^{-1} \hat{\mu} / 1' \hat{\Sigma}^{-1} 1$ the estimated mean of the minimum-variance portfolio (Kan and Zhou, 2007). Tu–Zhou: $w = \delta w_{\text{KZ3}} + (1 - \delta) 1/N$ (Tu and Zhou, 2011).

Cell. $D = I$ ($K = N$). In risk units, with Σ known, $\hat{\mu}_g \Sigma^{-1} 1$ becomes $u_0 u_0' \hat{b}$ with $u_0 = \Sigma^{-1/2} 1 / \|\Sigma^{-1/2} 1\|$, so the three-fund rule is

$$\gamma v = c [u_0 u_0' + \eta (I - u_0 u_0')] \hat{b} :$$

a soft reduction with two scalars, c on the level direction and $c\eta$ on the $N - 1$ cross-sectional directions. It is a reduction of $\hat{b}$, not an anchor: both components come from the forecast. Tu–Zhou add a fixed direction that does not depend on the forecast, $1/N$, which makes the rule affine in $\hat{b}$ (Corollary 2 with $x_0 = (1 - \delta)1/N$ and Q the three-fund map times δ).

Loss. For the three-fund rule, the general form of Theorem 2 with $\tilde{Q} = c[u_0 u_0' + \eta(I - u_0 u_0')]$, which separates into a level part and a cross-sectional part because the two projections are orthogonal: with $\theta_0^2 = (u_0' a)^2$ and $\theta_1^2 = \theta^2 - \theta_0^2$, $2\gamma \mathbb{E}[\text{loss}] = \theta^2 - 2\alpha_N (c\theta_0^2 + c\eta\theta_1^2) + \beta_N (c^2\theta_0^2 + c^2\eta^2\theta_1^2 + s^2(c^2 + c^2\eta^2(N - 1)))$. Minimizing over c and $c\eta$ separately gives $c^* = (\alpha_N / \beta_N) \theta_0^2 / (\theta_0^2 + s^2)$ on the level and $(c\eta)^* = (\alpha_N / \beta_N) \theta_1^2 / (\theta_1^2 + (N - 1)s^2)$ on the cross-section: the level, one dimension with the full signal of the minimum-variance direction, is shrunk little when its squared Sharpe ratio exceeds s^2 by a wide margin; the cross-section, $N - 1$ dimensions, is shrunk by the two-fund factor with $N - 1$ in place of N . These optimal values contain θ_0^2 and θ_1^2 and are oracle values; the published rules replace them by bias-corrected estimates. The decomposition above takes u_0 as known,

so that Q is fixed, and $\hat{\Sigma}$ only in the inverse; with u_0 estimated the exact expectation needs the further Wishart moments of [Kan and Zhou \(2007\)](#). For Tu–Zhou, Corollary 2: the anchor $u = (1 - \delta)\gamma\Sigma^{1/2}1/N$, whose direction has components on the level and on the cross-section, covers part of the signal that the shrinkage discards. Both rules stay inside the full basis with its price α_N, β_N .

3.4 The minimum-variance portfolio

Definition. $w \propto \hat{\Sigma}^{-1}1$: the forecast is replaced by a constant.

Cell. $D = I$ ($K = N$): the full inverse is needed. Q projects the forecast onto its level, $k = 1$; in risk units the direction is $u_0 = \Sigma^{-1/2}1/\|\Sigma^{-1/2}1\|$. In return units this projection is $Q\hat{\mu} = 1 \cdot (1'\Sigma^{-1}\hat{\mu})/(1'\Sigma^{-1}1)$, which depends on Σ ; with $\hat{\Sigma}$ in its place Q is data-dependent and Theorem 2 holds up to that (Remark 8). Length fixed, or the level estimated.

Loss. Theorem 2 with $K = N$, $k = 1$: $\theta^2 - (u'_0a)^2$ discarded, $(1 - 2\alpha_N + \beta_N)(u'_0a)^2$ inversion error, $\beta_N s^2$ noise. The cell is asymmetric: one forecast dimension, N covariance dimensions. It is the natural cell when the forecast has no cross-sectional content, and it pays the full covariance price for the one direction it trades.

3.5 The equal-weight portfolio

Definition. $w = 1/N$ ([DeMiguel et al., 2009b](#)).

Cell. $D = 1_N/N$, a single basis portfolio ($K = 1$); Q its level ($k = 1$); length fixed, or the portfolio's own sample mean over variance.

Loss. With a fixed length ℓ in risk units along $e = \Sigma^{1/2}1/\|\Sigma^{1/2}1\|$, Lemma 1 gives $2\gamma\text{loss} = \theta^2 - 2\ell e'a + \ell^2$, bounded below by $\theta^2(1 - \cos^2(e, a))$ and attaining it at $\ell = e'a$ (Proposition 8); the Sharpe ratio $\theta \cos(e, a)$ does not depend on the length. With the length estimated from the portfolio's own history, Theorem 2 with $K = k = 1$ adds $(1 - 2\alpha_1 + \beta_1)(e'a)^2 + \beta_1 s^2$, and $\alpha_1, \beta_1 \approx 1$. Nothing is inverted, nothing is amplified. The rule discards all signal except what the equal-weight direction carries and pays neither noise nor covariance price: the corner $K = k = 1$, the cell the theory prescribes when the equal-weight direction is the only one worth keeping and a basis large enough to hold the minimum-variance portfolio costs more than it adds. (When no direction at all passes the keep-or-kill test, the prescription is the risk-free asset, $k = 0$.)

3.6 The decile long–short sort

Definition. Rank the assets by the forecast; long the top decile, short the bottom decile, equal-weighted within each (Fama and French, 1993).

Cell. $D = d_{\text{sort}} = 1_{\text{top}}/n - 1_{\text{bot}}/n$, a single basis portfolio ($K = 1$), chosen by the ranking. Q : the sign ($k = 1$, magnitude fixed by convention). Length fixed. No covariance matrix is estimated.

Loss. With a fixed length ℓ along $d \propto \Sigma^{1/2}d_{\text{sort}}$, $2\gamma \text{ loss} = \theta^2 - 2\ell d'a + \ell^2$, bounded below by $\theta^2(1 - \cos^2(d, a))$ (Proposition 8). In the terms of Proposition 6 the sort uses the three-group partition {bottom, middle, top} that is optimal for a profile flat in the middle and stepping at the ends, but keeps only one of its three dimensions, top minus bottom, and discards the level and any asymmetry between the two ends; the tails are set to ± 1 instead of their estimated means. A basis of one portfolio inside a partition of three, which discards the middle of the ranking whatever the profile. Because $K = 1$ it pays no covariance price and treats the two tails as equally risky.

3.7 The parametric portfolio policy

Definition. $w = 1/N + \vartheta z/N$, z a standardized characteristic, ϑ maximizing the sample average utility of the policy over the window (Brandt et al., 2009).

Cell. $D = [1_N/N, z/N]$, two basis portfolios ($K = 2$): the sample utility of $w(\vartheta)$ is a quadratic in ϑ whose coefficients are the sample mean of the tilt portfolio, its variance and its covariance with $1/N$, i.e. the 2×2 sample moments of the basis. The weight on $1/N$ is fixed at one, so the estimated ϑ is the conditional optimum for the tilt given the benchmark, $\hat{\vartheta} = (\hat{m}_2 - \gamma \hat{V}_{12})/(\gamma \hat{V}_{22})$: a scalar inverse and an estimated hedge ratio $\hat{V}_{12}/\hat{V}_{22}$, not the inverse of the 2×2 matrix ($k = 1$, the slope of a line in z).

Loss. With V known the rule is affine (Corollary 2) with anchor $x_0 = (1, -V_{21}/V_{22})'$, the equal-weight portfolio hedged against the tilt, and $\tilde{Q}$ the orthogonal projection onto the tilt's risk direction ($k = 1$). Everything outside the span of the two portfolios is untradable, and the anchor covers the part of the tradable signal that the equal-weight portfolio carries net of the tilt. With $\hat{V}$ the variance of the tilt is estimated (constants $\alpha_1, \beta_1 \approx 1$), and the estimated hedge ratio adds a term of order $1/T$ times the anchor's residual risk that the corollary does not contain. The rule is a hard reduction to a line through the ranking with the level anchored at equal weights; its granularity is $k = 1$ whatever the characteristic can deliver.

3.8 Rank tiers

Definition. Rank the assets, form G groups, and hold the mean–variance position on the G equal-weighted group portfolios using their sample means and $G \times G$ sample covariance matrix over the window (Branger et al., 2019).

Cell. $D =$ the G group portfolios ($K = G$), chosen by the ranking; $Q = I$ ($k = G$): the group means are used as they are. Length a scalar with the constants of G .

Loss. At unit length, Theorem 2 with $K = k = G$: $\theta^2 - \|\Pi_D a\|^2$ untradable, nothing discarded inside, $(1 - 2\alpha_G + \beta_G)\|\Pi_D a\|^2$, $\beta_G G s^2$. Here both sides are reduced by the same partition, so the covariance price is paid on G dimensions and the untradable term is the within-group dispersion of the profile when the groups are exchangeable in covariance (Appendix A.1). The cell requires that an asset’s group membership in a past month be independent of the forecast error in that month; a forecast formed from the same returns violates this (Section 4). With G fixed the cell does not move. The number of groups is set by convention; Cattaneo et al. (2020) choose it from the data for inference on the return–characteristic relation, Section 4 for the investor’s loss.

3.9 Factor, shrinkage and principal-component covariance matrices

Three families reduce the covariance side and leave the forecast side alone.

Factor covariance. $\hat{\Sigma} = \hat{B}\hat{F}\hat{B}' + \hat{\Delta}$ with K factors. A hard reduction of Σ to K risk directions plus a diagonal; in the language of the risk ordering after Proposition 5 it keeps the high-variance directions and treats the rest as uncorrelated. It is not a restriction of the traded portfolios ($D = I$), so Theorem 2 does not apply verbatim; the covariance price falls to that of K factor dimensions plus N variances, and the forecast still enters with $k = N$.

Shrinkage. $\hat{\Sigma}_{\text{LW}} = \delta F + (1 - \delta)\hat{\Sigma}$ with a structured target F (Ledoit and Wolf, 2004). The soft reduction of the covariance side: every entry is pulled toward the target, none set to zero. It is to $\hat{\Sigma}$ what the Kan–Zhou scalar is to $\hat{\mu}$: $\hat{\Sigma}_{\text{LW}}^{-1}$ rotates the position, but k stays at N , no dimension of the forecast is dropped. Norm constraints on the weights act in the same way (DeMiguel et al., 2009a).

Principal components. $D =$ the top- K eigenvectors of Σ , in practice estimated on a separate sample (eigenvectors of $\hat{\Sigma}$ from the same window violate Assumption 1). A basis in the strict sense of Section 2.4, chosen by variance: the untradable signal is whatever lies in the dropped low-variance directions, which by the discussion after Proposition 5 can be much of $\|a\|^2$. The forecast side is free. Kozak et al. (2020) combine a truncation

of this kind with ridge shrinkage of the remaining components, a soft reduction inside a hard one (Remark 2).

None of the three moves with the forecast.

3.10 What the grid shows

Three observations follow from Table 1 read with Section 2.

First, the estimation-risk remedies (Kan–Zhou, Tu–Zhou, shrinkage) are soft: they keep every dimension of forecast and covariance matrix and scale. The practitioners’ rules ($1/N$, sort, tilt, tiers) are hard: they keep a few dimensions and discard the rest. Proposition 3 and Remark 3 say when hard is cheaper: when the signal is concentrated in few directions.

Second, the hard rules differ only in K , k and how the retained dimensions are chosen: one direction fixed by nothing ($1/N$), one fixed by the ranking (sort), one fixed by a characteristic (tilt), G fixed by the analyst (tiers). Each can be the best fixed cell for some profile, noise level and K/T , and none adapts to them.

Third, the two sides are chosen independently and the familiar rules combine them in every way: the minimum-variance portfolio keeps one forecast dimension inside a basis of N ; the sort keeps one inside a basis of one; the tiers keep G inside G ; the tilt keeps one inside two. Theorem 2 and Corollary 2 price every combination with a fixed basis and a fixed reduction; the covariance families of Section 3.9 lie outside. What no cell of the table does is set K or k from a measurement of the forecast; Section 4 takes that step. The result does not contradict the virtue of complexity in return prediction (Kelly et al., 2024), which concerns the number of parameters of the forecasting model; the price derived here is per dimension of the forecast that the position acts on, and a complex model may well deliver a forecast whose best use is coarse.

3.11 Connections to portfolio theory and asset pricing

The grid organizes two literatures that have mostly been read apart: the portfolio-choice literature on estimation risk, which supplies the rules of Table 1, and the asset-pricing literature on the structure of expected returns, which says where in the space of risk directions the signal is likely to sit. Both bear on the choice of the basis.

Portfolio theory. That the plug-in portfolio is dominated by its estimation error is the subject of Jobson and Korkie (1980), Best and Grauer (1991) and Britten-Jones (1999), and the mechanism, that the optimizer overweights whatever looks cheaply attractive in the sample, is Michaud’s error maximization (Michaud, 1989); Section 2.5 gives it

the form of a spread of sample eigenvalues and the price of Proposition 4. [Green and Hollifield \(1992\)](#) show that the mean–variance portfolio is concentrated when one factor dominates the covariance matrix: in the grid this is the level direction u_0 absorbing the inversion error of an N -dimensional basis, which the minimum-variance and three-fund cells pay in full. The Bayesian branch, [Jorion \(1986\)](#), [Pástor \(2000\)](#) and [Pástor and Stambaugh \(2000\)](#), shrinks the forecast toward a prior, in the last two toward the span of an asset-pricing model: a soft reduction whose target is a low-dimensional subspace, i.e. a soft version of a basis. [Garlappi et al. \(2007\)](#) obtain the same shrinkage from ambiguity aversion, and [DeMiguel et al. \(2009a\)](#) from norm constraints on the weights. All of these keep $k = N$. The hard side of the literature is the practitioners’ side: [DeMiguel et al. \(2009b\)](#) on $1/N$, [Brandt et al. \(2009\)](#) on the parametric tilt, and the sorting convention since [Fama and French \(1993\)](#). [Branger et al. \(2019\)](#) are the closest predecessor: they sort assets into equally weighted groups, optimize over the groups and derive the optimal group size, which rises with estimation risk; their strategy is the rank-tiers cell and their group size the dimension $K = G$ of Corollary 4. A single group size makes two choices at once, the basis in which the covariance matrix is estimated and the number of forecast dimensions acted on; the present paper separates them, and Theorem 2 prices each. On the forecast side, the ranking literature supplies the reductions of Section 2.6: [Guo and Wachter \(2026\)](#) show for a single asset that even the sign of a forecast carries a premium, the rank-weighted long–short portfolios of [Asness et al. \(2013\)](#) are the line at full resolution without a covariance matrix, and the information coefficient that sets G is the practitioner’s measure of ranking skill since [Grinold \(1989\)](#).

Asset pricing: where the signal sits. Whether a basis ordered by risk is a good one depends on where a has its length. [Kozak et al. \(2018\)](#) argue that in the absence of near-arbitrage the stochastic discount factor must load on the dominant principal components of returns: a large Sharpe ratio in a low-variance direction would be a near-arbitrage. In the terms of Proposition 5, most of $\|a\|^2$ should then be tradable with a few high-variance portfolios, and whatever the plug-in finds in the small eigenvalues is estimation noise, which is what the spread of Section 2.5 produces mechanically. [Kozak et al. \(2020\)](#) turn this into an estimator by shrinking the loadings on low-variance components toward zero. The empirical literature on characteristics points the other way: [Daniel et al. \(2020\)](#) show that characteristic-sorted portfolios hedged against their factor exposures have higher Sharpe ratios than the unhedged ones, so that the signal the ranking carries survives, and is concentrated, in directions of low variance; [Lettau and Pelger \(2020\)](#) choose principal components by a criterion that weights the mean as well as the variance, and [Kelly et al. \(2019\)](#) let characteristics instrument the loadings. The two views are not in conflict about the facts, they are about different bases: the risk ordering (eigenvectors) and the forecast ordering (ranks) agree on where the large-variance signal is and disagree on the rest. This

is why the grid keeps the two sides separate and why Proposition 5 prices a mixed basis: a few risk portfolios make the dominant directions tradable at a low covariance price, and a few rank portfolios make the hedged characteristic signal tradable at the price of their own dimension. It is also why the group covariance matrix of the tiers cell, which estimates risk between rank groups only, can beat both the full sample covariance matrix and no covariance matrix at all: it trades the hedged directions without inverting the dominant ones.

4 The decision rule

Every cell of Table 1 fixes its dimensions once. This section derives the cell that sets them from the data: a basis of G rank groups, a line through their means, a covariance matrix between the groups, and a length, with G read off the forecast's measured ranking ability. Section 4.1 states the cell and the two assumptions under which Theorem 2 applies to a basis that is rebuilt every month. Section 4.2 derives the number of groups from the two-sided loss, which is the one place where the theory and the rule have to be reconciled: the number of groups that Corollary 3 gives for steps and the number that the two-sided loss gives for the line differ, and we show by how much. Section 4.3 lists the six steps of the rule with the result each one rests on.

4.1 The cell and two assumptions

At each formation date the investor has N assets, their excess returns over the past T months, and a forecast score for each asset in each of those months and today. He ranks the assets by today's score, cuts the ranking into G contiguous groups of equal size, and trades the G equal-weighted group portfolios: D is the $N \times G$ matrix of group indicators divided by group size, $K = G$. For the history of the basis he uses, for each past month, the return of the equal-weighted portfolio of the assets that were in group g *in that month*, so that the G series y_t are the returns of managed portfolios whose membership changes with the ranking. From these series he forms $\hat{m}$ and $\hat{V}$, replaces $\hat{m}$ by the fitted values of a regression of the G group means on a constant and the group rank ($Q = Q_{\text{line}}$, reduction (ii) of Section 2.6), and holds

$$w = Dx, \quad x = \frac{\hat{\kappa}}{\gamma} \hat{V}^{-1} Q_{\text{line}} \hat{m}, \quad (39)$$

with $\hat{\kappa}$ the feasible scalar of Appendix A.2 with constants α_G, β_G and $k = 2$, or a volatility target when the object of interest is the Sharpe ratio. Every asset receives its group's weight divided by the number of assets in the group. With $G = 1$ the basis is the equal-weight portfolio and the position is $1/N$ at the chosen length.

Two features of this cell are not covered by Assumption 1 as stated, and we make the assumptions under which they are.

Assumption 2 (Managed group portfolios). The monthly returns of the G managed group portfolios, each formed from the assets in group g at the start of the month, are i.i.d. normal with mean m and covariance matrix V , and today's group portfolios have the same m and V .

The basis of Theorem 2 is a fixed matrix with a history of its own returns. A rank-group basis is a sequence of matrices, one per month, and its “history” is the record of what the group- g portfolio earned when its membership was that month's. Assumption 2 says that this record is a sample from the distribution of today's group portfolio: group g is a stable object (“the top fifth by the forecast”) even though its members change. This is the sense in which a sorted portfolio has a mean and a variance in the asset-pricing literature, and it is the assumption under which the group means and the group covariance matrix are estimable at all.

Assumption 3 (Membership and forecast error). The group membership of an asset in a past month is independent of the forecast error of that month's returns, i.e. of $\hat{m} - m$ for the group series.

Assumption 3 is what makes the ranks, and hence D and Q_{line} , fixed in the sense of Assumption 1 and Remark 8. For the cell as stated it holds by construction under i.i.d. returns, also when the score is formed from past returns: the membership in month t is set by the score available at the start of month t , which uses only earlier returns, so the group's return in month t is not selected on. It fails when the history is computed with *today's* members: then the top group is the group of assets whose past returns happened to be high, its measured mean carries the selection, and the group means are biased upward; this is why step 3 below uses the membership at each date. Under both assumptions the cell is priced by Theorem 2 with $K = G$ and $k = 2$, and its basis is the one that Proposition 10 relates to a group-mean projection.

4.2 The number of groups

The number of groups is the only input the forecast sets, and Section 2.6 offers two formulas for it. Corollary 3 with Proposition 7 gives, for *steps* with a known covariance matrix,

$$G_{\text{steps}}^* = \left(\frac{2K\tau^2}{s^2} \right)^{1/3} = \left(\frac{2KT\rho^2}{1-\rho^2} \right)^{1/3}, \quad (40)$$

balancing the signal lost inside groups, $K\tau^2/G^2$, against the noise Gs^2 of G estimated group means. The rule uses the line, not the steps, and estimates the covariance matrix in the group space. Both change the count.

Proposition 9 (Number of groups for the rule). *Let the setting of Proposition 7 hold (uncorrelated basis, linear profile $\tilde{a}_i = \tilde{a} + \tau z_i$, isotropic error), let $S_{\text{lev}} = K\tilde{a}^2$ and $S_{\text{rank}} = K\tau^2$ be the squared Sharpe ratios of the level and of the ranking, and $S_{\infty} = S_{\text{lev}} + S_{\text{rank}}$ their sum, the signal tradable with every asset. For the cell (39) at unit length, under Assumptions 1, 2 and 3, to first order in G/T ,*

$$2\gamma \mathbb{E}[\text{loss}](G) \approx \|a\|^2 - S_{\infty} + \frac{S_{\text{rank}}}{G^2} + \frac{G}{T} S_{\infty} + \frac{2}{T}, \quad (41)$$

minimized at

$$G_{\text{line}}^* = \left(\frac{2T S_{\text{rank}}}{S_{\infty}} \right)^{1/3} = \left(2T \cdot \frac{\tau^2}{\tilde{a}^2 + \tau^2} \right)^{1/3} \leq (2T)^{1/3}. \quad (42)$$

The same G_{line}^* obtains with the optimal scalar in place of unit length.

Proof. Apply Theorem 2 with $K = G$ and $\tilde{Q}$ the line, an orthogonal projection of rank $k = 2$ in the risk units of the group basis (the regression is fitted in the metric of V ; with equal groups and an uncorrelated basis of equal variances the group portfolios have equal variances, and the fit coincides with ordinary least squares). The four terms are: the untradable signal $\|a\|^2 - S_G$ with S_G the signal tradable with G equal groups, which by Corollary 3 and Proposition 10 is $S_{\infty} - K\tau^2 \frac{(K/G)^2 - 1}{K^2 - 1} \approx S_{\infty} - S_{\text{rank}}/G^2$; the signal discarded inside, $S_G - \|\tilde{Q}\tilde{a}_G\|^2$, which is zero because the true group means in risk units are exactly linear in the group rank when the profile is linear and the groups are equal (the mean rank of group g is equally spaced in g); the inversion error $(1 - 2\alpha_G + \beta_G) S_G$; and the noise $\beta_G \cdot 2s^2$. Collecting with $\varphi_G = 2\alpha_G - \beta_G$ as in (23), $2\gamma \mathbb{E}[\text{loss}] = \|a\|^2 - \varphi_G S_G + 2\beta_G s^2$. Expand for G/T small as in the proof of Corollary 4, $\varphi_G \approx 1 - G/T$ and $\beta_G \approx 1 + 3G/T$:

$$\|a\|^2 - \left(1 - \frac{G}{T}\right) \left(S_{\infty} - \frac{S_{\text{rank}}}{G^2}\right) + \frac{2}{T} \left(1 + \frac{3G}{T}\right) = \|a\|^2 - S_{\infty} + \frac{S_{\text{rank}}}{G^2} + \frac{G}{T} S_{\infty} + \frac{2}{T} - \frac{S_{\text{rank}}}{GT} + \frac{6G}{T^2}.$$

The last two terms are of smaller order than S_{rank}/G^2 and GS_{∞}/T respectively (their ratios are G/T and $6/(TS_{\infty})$) and are dropped, which gives (41). The function of G is convex with derivative $-2S_{\text{rank}}/G^3 + S_{\infty}/T$, zero at $G^3 = 2TS_{\text{rank}}/S_{\infty}$. With $S_{\text{rank}} = K\tau^2$ and $S_{\infty} = K(\tilde{a}^2 + \tau^2)$ the K cancels. The bound follows from $S_{\text{rank}} \leq S_{\infty}$. With the optimal scalar, Appendix A.2 gives the loss as $\|a\|^2 - (\alpha_G^2/\beta_G) S_G^2/(S_G + 2s^2)$, and $\alpha_G^2/\beta_G \approx 1 - (G + 1)/T$; the coefficient of G/T is now $S_G^2/(S_G + 2s^2)$ instead of S_G , so the first-order condition is unchanged up to the factor $(S_{\infty} + 4s^2)/(S_{\infty} + 2s^2)$, which is close to one when $2s^2$ is small relative to S_{∞} . $\square$

Three things follow. First, with the line the forecast side no longer prices the number of groups: its noise is $2s^2$ whatever G , and the term that rises with G is the covariance price GS_{∞}/T , the first-order inversion error on the signal already tradable. The number

of groups is set by the covariance side, which is the content of Corollary 4 applied to this cell. Second, there is a ceiling: $G_{\text{line}}^* \leq (2T)^{1/3}$, about six groups with ten years of monthly data, whatever the forecast. No ranking ability justifies finer groups when the covariance matrix between them is estimated from the same window; a finer ranking is used through the line, not through more groups. Third, the count rises with the forecast's ranking ability through the share $S_{\text{rank}}/S_{\infty}$ of the tradable signal that the ranking carries, from one group when the ranking carries nothing to the ceiling when it carries everything, and it does not depend on K : with the line, a hundred assets and thirty assets want the same number of groups for the same forecast quality.

Reconciling the two formulas. The rule's step 2 uses (40), which has only observable inputs (K , T and the measured ρ), while (42) needs the share $S_{\text{rank}}/S_{\infty}$, a ratio of two squared Sharpe ratios that is estimated far less precisely than ρ . The two differ by the factor

$$\frac{G_{\text{line}}^*}{G_{\text{steps}}^*} = \left(\frac{1}{S_{\infty}} \right)^{1/3} = \left(\frac{1}{K(\bar{a}^2 + \tau^2)} \right)^{1/3}, \quad (43)$$

which exceeds one when the signal tradable with all K assets has a squared monthly Sharpe ratio below one, the empirically relevant case: the line count is then the larger of the two. The rule nevertheless keeps (40), for two reasons. Its inputs are measured with a standard error of about $1/\sqrt{KT}$, whereas $S_{\text{rank}}/S_{\infty}$ is a ratio of two noisy squared Sharpe ratios; and (40) rounds to one group whenever ρ is at its estimation noise, which is the case that matters most. The uncorrelated basis of Proposition 9 does not describe individual stocks well, so we do not evaluate (42) numerically here; which of the two counts does better out of sample is an empirical question. The granularity curve of the application has its optimum at four groups for the machine-learning forecast, where (40) also gives about four.

4.3 The six steps and what each rests on

1. *Rank* the N assets by today's forecast. The ranking is the only use of the forecast's values; its magnitudes are not used (Section 2.6, reductions (i)–(ii)). Under Assumption 3 the ranks are fixed in the sense of Theorem 2.
2. *Number of groups.* $G = (2NT\rho^2/(1 - \rho^2))^{1/3}$ from the trailing rank correlation ρ of the forecast with the next month's return, measured on the months before the holding month and rounded to the nearest of $\{1, 2, 3, 5, 10\}$ (Corollary 3 and Proposition 7). $\rho \leq 0$ gives one group. Ten groups, above the ceiling $(2T)^{1/3}$ of Proposition 9, require ρ above about 0.13 at $N = 100$ and $T = 120$ and do not occur in the application.

3. *Group histories.* For each of the past T months, the equal-weighted return of the assets that were in group g in that month: the G series of the managed basis (Assumption 2). Equal weights within a group are the mean-variance position when the forecast cannot distinguish the members and the covariance inside a group is not estimated (Proposition 10).
4. *Magnitudes.* A line through the G group means (Q_{line} , rank two): noise $2s^2$ instead of Gs^2 , no signal lost when the profile is linear, and the number of groups then set by the covariance side (Proposition 9).
5. *Weights.* Mean-variance on the G groups with their $G \times G$ sample covariance matrix: the basis of Section 2.4 with $K = G$, priced by Theorem 2 with constants α_G, β_G near one for $G \leq 6$ and $T = 120$.
6. *Length.* The feasible scalar of Appendix A.2 with $K = G$ and $k = 2$ for the certainty-equivalent investor, or a volatility target for the Sharpe-ratio investor (Proposition 8).

Only step 2 depends on the forecast. A forecast that cannot rank gives one group, and the rule is $1/N$: the corner of Table 1 that the theory prescribes when no cross-sectional direction passes the keep-or-kill test and the covariance matrix is not worth estimating. A forecast that can rank gives several groups, and the rule is the tiers cell with a line and a group covariance matrix: an interior point of the grid chosen ex ante, which the application compares with every fixed cell.

4.4 An illustration

We apply the rule to large and mega capitalization stocks in CRSP, from December 1989 to November 2023 (408 months, $T = 120$). Each December we draw $N = 100$ stocks; the results average over 100 such draws. We use six forecasts of increasing quality: the historical mean, momentum, a two-characteristic composite, a fifteen-characteristic composite, a partial-least-squares forecast and a machine-learning forecast (rank information coefficient about 0.06). Table 2 reports annualized out-of-sample Sharpe ratios.

Three patterns follow the theory. In the application the rule never falls below $1/N$: when the measured information coefficient is close to zero, $G = 1$ in most months and the rule is close to $1/N$. (Because the coefficient is measured with noise, a forecast with no ranking ability still yields $G = 2$ in some months, so the rule can fall slightly below $1/N$; it does not in Table 2.) Its Sharpe ratio rises with forecast quality, because G rises with the information coefficient. The fixed cells are right at one point of the ladder only: the decile sort keeps one dimension whatever the forecast and loses heavily on weak forecasts, the

Table 2: Annualized out-of-sample Sharpe ratios, 100 large CRSP stocks redrawn each December, 1989–2023, average over 100 draws. The plug-in portfolio earns about zero for every forecast; $1/N$ does not use the forecast and earns 0.56. Preliminary; confidence intervals and ablations follow in the next version.

Forecast	Rule (Section 4)	Decile sort	Parametric tilt
Historical mean	0.57	−0.12	0.33
Momentum	0.58	0.09	0.41
Two-characteristic composite	0.62	0.24	0.46
Fifteen-characteristic composite	0.73	0.29	0.55
Partial least squares	0.92	0.59	0.88
Machine learning	1.22	0.99	1.23

parametric tilt matches the rule on the strongest forecast and falls behind on the weaker ones, and the plug-in, which keeps all N dimensions on both sides, earns nothing.

5 Conclusion

Using less of a forecast is not a heuristic but a priced choice. Each dimension of the forecast an investor acts on costs the variance of an estimated Sharpe ratio, each dimension of the covariance matrix he estimates costs a Wishart factor, and the familiar portfolio rules are fixed settings of these two choices. The setting that adapts to the forecast follows from the same prices: the number of rank groups grows with the cube root of $NT\rho^2/(1-\rho^2)$, and the rule falls back to $1/N$ when the forecast has nothing to say. Less is more exactly when the signal a dimension adds is smaller than its price.

References

- Asness, C. S., Moskowitz, T. J., and Pedersen, L. H. (2013). Value and momentum everywhere. *The Journal of Finance*, 68(3):929–985.
- Best, M. J. and Grauer, R. R. (1991). On the sensitivity of mean-variance-efficient portfolios to changes in asset means: Some analytical and computational results. *The Review of Financial Studies*, 4(2):315–342.
- Brandt, M. W., Santa-Clara, P., and Valkanov, R. (2009). Parametric Portfolio Policies: Exploiting Characteristics in the Cross-Section of Equity Returns. *The Review of Financial Studies*, 22(9):3411–3447.
- Branger, N., Lučivjanská, K., and Weissensteiner, A. (2019). Optimal granularity for portfolio choice. *Journal of Empirical Finance*, 50:125–146.
- Britten-Jones, M. (1999). The sampling error in estimates of mean-variance efficient portfolio weights. *The Journal of Finance*, 54(2):655–671.

- Cattaneo, M. D., Crump, R. K., Farrell, M. H., and Schaumburg, E. (2020). Characteristic-sorted portfolios: Estimation and inference. *Review of Economics and Statistics*, 102(3):531–551.
- Chan, L. K. C., Karceski, J., and Lakonishok, J. (1999). On portfolio optimization: Forecasting covariances and choosing the risk model. *The Review of Financial Studies*, 12(5):937–974.
- Daniel, K., Mota, L., Rottke, S., and Santos, T. (2020). The cross-section of risk and returns. *The Review of Financial Studies*, 33(5):1927–1979.
- DeMiguel, V., Garlappi, L., Nogales, F. J., and Uppal, R. (2009a). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science*, 55(5):798–812.
- DeMiguel, V., Garlappi, L., and Uppal, R. (2009b). Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy? *The Review of Financial Studies*, 22(5):1915–1953.
- Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.
- Garlappi, L., Uppal, R., and Wang, T. (2007). Portfolio Selection with Parameter and Model Uncertainty: A Multi-Prior Approach. *Review of Financial Studies*, 20(1):41–81.
- Green, R. C. and Hollifield, B. (1992). When Will Mean-Variance Efficient Portfolios Be Well Diversified? *The Journal of Finance*, 47(5):1785–1809.
- Grinold, R. C. (1989). The fundamental law of active management. *The Journal of Portfolio Management*, 15(3):30–37.
- Gu, S., Kelly, B., and Xiu, D. (2020). Empirical Asset Pricing via Machine Learning. *The Review of Financial Studies*, 33(5):2223–2273.
- Guo, H. and Wachter, J. A. (2026). Forecast-agnostic portfolios. Working paper, 9 February 2026.
- Haff, L. R. (1979). An identity for the Wishart distribution with applications. *Journal of Multivariate Analysis*, 9(4):531–544.
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, NY, 2nd edition.
- Jagannathan, R. and Ma, T. (2003). Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps. *The Journal of Finance*, 58(4):1651–1683.

- James, W. and Stein, C. (1961). Estimation with quadratic loss. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 361–379. University of California Press.
- Jobson, J. D. and Korkie, B. (1980). Estimation for Markowitz Efficient Portfolios. *Journal of the American Statistical Association*, 75(371):544–554.
- Jobson, J. D. and Korkie, B. M. (1981). Performance hypothesis testing with the Sharpe and Treynor measures. *The Journal of Finance*, 36(4):889–908.
- Jorion, P. (1986). Bayes-Stein Estimation for Portfolio Analysis. *The Journal of Financial and Quantitative Analysis*, 21(3):279–292.
- Kan, R. and Zhou, G. (2007). Optimal Portfolio Choice with Parameter Uncertainty. *The Journal of Financial and Quantitative Analysis*, 42(3):621–656.
- Kelly, B., Malamud, S., and Zhou, K. (2024). The virtue of complexity in return prediction. *The Journal of Finance*, 79(1):459–503.
- Kelly, B. T., Pruitt, S., and Su, Y. (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, 134(3):501–524.
- Klein, R. W. and Bawa, V. S. (1976). The effect of estimation risk on optimal portfolio choice. *Journal of Financial Economics*, 3(3):215–231.
- Kozak, S., Nagel, S., and Santosh, S. (2018). Interpreting factor models. *The Journal of Finance*, 73(3):1183–1223.
- Kozak, S., Nagel, S., and Santosh, S. (2020). Shrinking the cross-section. *Journal of Financial Economics*, 135(2):271–292.
- Ledoit, O. and Wolf, M. (2004). Honey, I Shrunk the Sample Covariance Matrix. *The Journal of Portfolio Management*, 30(4):110–119.
- Ledoit, O. and Wolf, M. (2017). Nonlinear Shrinkage of the Covariance Matrix for Portfolio Selection: Markowitz Meets Goldilocks. *The Review of Financial Studies*, 30(12):4349–4388.
- Lettau, M. and Pelger, M. (2020). Estimating latent asset-pricing factors. *Journal of Econometrics*, 218(1):1–31.
- Lo, A. W. (2002). The statistics of Sharpe ratios. *Financial Analysts Journal*, 58(4):36–52.
- Lopez de Prado, M. (2016). Building Diversified Portfolios that Outperform Out of Sample. *The Journal of Portfolio Management*, 42(4):59–69.

- Marčenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4):457–483.
- Markowitz, H. (1952). Portfolio Selection*. *The Journal of Finance*, 7(1):77–91.
- Michaud, R. O. (1989). The Markowitz Optimization Enigma: Is ‘Optimized’ Optimal? *Financial Analysts Journal*, 45(1):31–42.
- Pástor, L. (2000). Portfolio Selection and Asset Pricing Models. *The Journal of Finance*, 55(1):179–223.
- Pástor, L. and Stambaugh, R. F. (2000). Comparing asset pricing models: An investment perspective. *Journal of Financial Economics*, 56(3):335–381.
- Pedersen, L. H., Babu, A., and Levine, A. (2021). Enhanced portfolio optimization. *Financial Analysts Journal*, 77(2):124–151.
- Salcher, L., Stöckl, S., and Hanke, M. (2026). Lost in Translation? Risk-Adjusting RMSE for Economic Forecast Performance. *Journal of Forecasting*, n/a(n/a):1–19.
- Tu, J. and Zhou, G. (2011). Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies. *Journal of Financial Economics*, 99(1):204–215.

A Appendix to Section 2

A.1 When group portfolios are a group-mean projection

Let $\mathcal{G} = \{g_1, \dots, g_G\}$ be a partition of the N assets, 1_g the indicator of group g , $U = \text{span}\{1_{g_1}, \dots, 1_{g_G}\}$, $P_{\mathcal{G}}$ the orthogonal projection onto U (group averaging), and $D = [1_{g_1}, \dots, 1_{g_G}]$ the basis of equal-weighted group portfolios. Two objects appear in the text that coincide only under a condition on Σ : the restriction of the action space to D , whose untradable term is $\|a\|^2 - \|\Pi_D a\|^2$ with Π_D the projection onto $\text{span}(\Sigma^{1/2} D)$ (Lemma 2), and the group-mean reduction of the forecast, whose discarded term is the within-group dispersion $\|(I - P_{\mathcal{G}})a\|^2$ (Proposition 6).

Proposition 10 (Group condition). *The following are equivalent:*

- (a) U is invariant under Σ : $\Sigma u \in U$ for every $u \in U$;
- (b) $\Sigma 1_g \in U$ for every group g , i.e. for every pair of groups (g, h) the row sums $\sum_{j \in g} \Sigma_{ij}$ take the same value for all $i \in h$;
- (c) $P_{\mathcal{G}} \Sigma = \Sigma P_{\mathcal{G}}$.

Under these conditions $\text{span}(\Sigma^{1/2}D) = U$, so that $\Pi_D = P_G$ and

$$\|a\|^2 - \|\Pi_D a\|^2 = \|(I - P_G)a\|^2 = \sum_g \sum_{i \in g} (a_i - \bar{a}_g)^2; \quad (44)$$

moreover the group-mean reduction in return units commutes with Σ and is therefore an orthogonal projection in risk units (Remark 7). A sufficient condition is that Σ be block-exchangeable: equal variances v_g and equal covariances c_{gg} within each group, and a constant covariance c_{gh} between any two groups.

Proof. (a) $\Leftrightarrow$ (b): the 1_g span U , so $\Sigma U \subseteq U$ if and only if each $\Sigma 1_g$ lies in U ; and a vector lies in U if and only if it is constant within each group, which for $\Sigma 1_g$ is the statement about row sums.

(a) $\Rightarrow$ (c): if U is invariant under the symmetric matrix Σ , so is $U^\perp$: for $x \in U^\perp$ and $u \in U$, $\langle \Sigma x, u \rangle = \langle x, \Sigma u \rangle = 0$. Write any y as $y = P_G y + (I - P_G)y$. Then $\Sigma P_G y \in U$ and $\Sigma(I - P_G)y \in U^\perp$, so $P_G \Sigma y = \Sigma P_G y$. (c) $\Rightarrow$ (a): for $u \in U$, $\Sigma u = \Sigma P_G u = P_G \Sigma u \in U$.

Invariance passes to $\Sigma^{1/2}$: an invariant subspace of a symmetric positive-definite matrix is spanned by some of its eigenvectors, and $\Sigma^{1/2}$ has the same eigenvectors. Since $\Sigma^{1/2}$ is invertible, $\Sigma^{1/2}U$ has the dimension of U and lies in U , hence equals U . The columns of D span U , so $\text{span}(\Sigma^{1/2}D) = \Sigma^{1/2}U = U$ and $\Pi_D = P_G$. Equation (44) follows from Lemma 2 and the expansion of $\|(I - P_G)a\|^2$ in Proposition 6. For the commutation statement note that P_G is the group averaging matrix Q of Remark 7, and (c) is exactly “ Q commutes with V ” for $D = I$.

Block exchangeability: for $i \in h$, $(\Sigma 1_g)_i = \sum_{j \in g} \Sigma_{ij}$ equals $n_g c_{hg}$ if $h \neq g$ and $v_g + (n_g - 1)c_{gg}$ if $h = g$; either way it does not depend on which $i \in h$ is taken, so (b) holds. $\square$

Without the condition, the two objects differ. The investor who holds equal-weighted group portfolios faces the untradable term $\|a\|^2 - \|\Pi_D a\|^2$, which is what Theorem 2 prices; the within-group dispersion is then an approximation to it, exact when the covariance structure respects the groups. The same condition, applied to V and groups of basis portfolios, makes the group-mean reduction inside a basis orthogonal in risk units, so that its noise is Gs^2 rather than more (Remark 7).

A.2 The scalar with an estimated covariance matrix

Let the investor hold $x = \kappa \hat{V}^{-1} Q \hat{m} / \gamma$ with a fixed Q whose risk-unit image $\tilde{Q}$ is an orthogonal projection of rank k , and write $S_Q = \|\tilde{Q} \tilde{a}\|^2$ for the signal the reduction retains.

Proposition 11 (Optimal and feasible scalar). *Under Assumption 1,*

$$2\gamma \mathbb{E}[\text{loss}](\kappa) = (\|a\|^2 - \|\tilde{a}\|^2) + \|\tilde{a}\|^2 - 2\alpha_K \kappa S_Q + \beta_K \kappa^2 (S_Q + ks^2), \quad (45)$$

minimized at

$$\kappa^* = \frac{\alpha_K}{\beta_K} \cdot \frac{S_Q}{S_Q + ks^2}, \quad 2\gamma \mathbb{E}[\text{loss}](\kappa^*) = \|a\|^2 - \frac{\alpha_K^2}{\beta_K} \cdot \frac{S_Q^2}{S_Q + ks^2}. \quad (46)$$

κ^* contains the unknown S_Q and is an oracle value. The in-sample squared Sharpe ratio of the reduced forecast, $\hat{\psi} = \hat{m}'Q'\hat{V}^{-1}Q\hat{m}$, has $\mathbb{E}[\hat{\psi}] = \alpha_K(S_Q + ks^2)$, so that $\hat{S}_Q = \hat{\psi}/\alpha_K - ks^2$ is unbiased for S_Q and the feasible scalar is

$$\hat{\kappa} = \frac{\alpha_K}{\beta_K} \cdot \frac{(\hat{\psi} - \alpha_K ks^2)^+}{\hat{\psi}}. \quad (47)$$

With $D = I$ and $Q = I$ this is the feasible two-fund rule of [Kan and Zhou \(2007\)](#) up to the normalization of $\hat{\Sigma}$; with V known ($\alpha_K = \beta_K = 1$) it is (38).

Proof. Replace $\tilde{Q}$ by $\kappa\tilde{Q}$ in the general form of Theorem 2: $\tilde{a}'(\kappa\tilde{Q})\tilde{a} = \kappa S_Q$, $\|\kappa\tilde{Q}\tilde{a}\|^2 = \kappa^2 S_Q$ and $\|\kappa\tilde{Q}\|_F^2 = \kappa^2 k$, which gives (45). It is a convex quadratic in κ with derivative $-2\alpha_K S_Q + 2\beta_K \kappa (S_Q + ks^2)$, zero at κ^* ; substituting, $-2\alpha_K \kappa^* S_Q + \beta_K (\kappa^*)^2 (S_Q + ks^2) = -\alpha_K \kappa^* S_Q = -(\alpha_K^2/\beta_K) S_Q^2/(S_Q + ks^2)$. For $\hat{\psi}$: by independence of $\hat{m}$ and $\hat{V}$, $\mathbb{E}[\hat{\psi}] = \alpha_K \mathbb{E}[\hat{m}'Q'V^{-1}Q\hat{m}] = \alpha_K (\|\tilde{Q}\tilde{a}\|^2 + s^2\|\tilde{Q}\|_F^2) = \alpha_K (S_Q + ks^2)$, the computation in the proof of Theorem 2. Inserting $\hat{S}_Q$ for S_Q in κ^* and simplifying $\hat{S}_Q/(\hat{S}_Q + ks^2) = (\hat{\psi}/\alpha_K - ks^2)/(\hat{\psi}/\alpha_K)$ gives (47); the positive part handles $\hat{\psi} < \alpha_K ks^2$, where the estimated signal is negative and the investor holds the risk-free asset. $\square$

Three comments. The feasible scalar inserts an unbiased estimate into a ratio, so it is not itself unbiased for κ^* ; [Kan and Zhou \(2007\)](#) show for $D = I$, $Q = I$ that the loss from this is small, and the argument carries over because only $\hat{\psi}$ enters. The scalar is a data-dependent reduction in the sense of Remark 8, and its own estimation cost is not in (46). And the subtraction $\alpha_K ks^2$ is the bias correction of Section 2.7 with the Wishart factor: the in-sample squared Sharpe ratio overstates the true one by the retained noise ks^2 , and the inversion inflates both by α_K .