Appendix to “Portfolio Construction by Mitigating Error Amplification: The Bounded-Noise Portfolio”

Long Zhao, Deepayan Chakrabarti, and Kumar Muthuraman
McCombs School of Business, University of Texas, Austin, TX, USA,
long.zhao@mccombs.utexas.edu,
deepayan.chakrabarti@mccombs.utexas.edu,
kumar.muthuraman@mccombs.utexas.edu

Appendix A: Absolute Versus Relative Differences in Eigenvalues

Lemma 3.2 shows that the bottom eigenvectors could not be well estimated because they are not well separated (i.e., the absolute differences between their eigenvalues are small). This observation might suggest that the corresponding eigenvectors are almost interchangeable and that these errors have a limited effect on the performance of the aggressive noise-only portfolio. However, this intuition is false, because the optimal portfolio depends on the *relative* differences between eigenvalues, which can still be large. We demonstrate this understanding with some examples.

Example 1. Suppose the eigenvalues vary as a (heavy-tailed) power-law, with $\lambda_i = \xi \cdot \beta^{p-i-1}$ for some $\xi > 0$ and $\beta > 1$. The absolute difference between consecutive eigenvalues is $\lambda_i - \lambda_{i+1} = \xi \cdot \beta^{p-i-1} \cdot (1 - 1/\beta)$, which decreases with increasing i and is at most ξ for the last two eigenvalues. Thus, with a large enough β and a small enough ξ, every consecutive pair of eigenvectors is well separated, except for the bottom two eigenvectors. Under these conditions, for some number of samples n, we can expect the top $p-2$ eigenvectors to be well estimated, and the last two to be poorly estimated. Let us also assume for simplicity that $\boldsymbol{v}_{p-1}'\mathbf{1} = \boldsymbol{v}_p'\mathbf{1} = \rho \neq 0$.

In general, the bottom eigenvalues are poorly estimated, as in Figure 1. However, let us consider the best-case scenario for estimation: Suppose that the top $p-2$ eigenvectors are estimated perfectly, as are all eigenvalues. Let us take $\hat{\mathcal{S}}$ to be the span of the first $p-2$ sample eigenvectors. Because these eigenvectors are perfectly estimated, we have $\hat{\mathcal{S}} = \mathcal{S}$. Let $\hat{\mathcal{N}}$ and $\mathcal{N}$ denote the spans of the last two sample eigenvectors and true eigenvectors, respectively. Note that $\hat{\mathcal{N}} = \mathcal{N}$ because each is simply the space orthogonal to $\hat{\mathcal{S}} = \mathcal{S}$. Thus, the only error is in the orientation of the bottom two eigenvectors, $\hat{\boldsymbol{v}}_{p-1}$ and $\hat{\boldsymbol{v}}_p$.

In other words,

$$\hat{\boldsymbol{v}}_{p-1} = \boldsymbol{v}_{p-1} \cdot \cos\theta - \boldsymbol{v}_p \cdot \sin\theta \qquad\qquad \hat{\boldsymbol{v}}_p = \boldsymbol{v}_{p-1} \cdot \sin\theta + \boldsymbol{v}_p \cdot \cos\theta$$

for some random angle θ. Because we can always reverse the direction of these four eigenvectors without loss of generality, we confine θ to $[0, \pi]$. Then, we can show:

$$RV(\boldsymbol{w}_N^*) = \frac{\xi}{\rho^2 \cdot (\beta + 1)} \tag{1}$$

$$RV(\hat{\boldsymbol{w}}_N^*) \approx EV(\hat{\boldsymbol{w}}_N^*) \cdot \left(1 + (\beta - 1) \cdot (\sin\theta)^2\right) \quad \text{for } \beta \gg 1 \tag{2}$$

$$RV(\hat{\boldsymbol{w}}_N^*) \approx RV(\boldsymbol{w}_N^*) \cdot \frac{\beta \cdot (\sin\theta)^2}{(\cos\theta + \sin\theta)^2} \quad \text{for } \beta \gg 1. \tag{3}$$

The approximations hold when $\theta \neq \frac{3\pi}{4}$, which is true with probability 1. Thus, the aggressive noise-only portfolio is considered to be a far better portfolio than it actually is ($EV(\hat{\boldsymbol{w}}_N^*) \ll RV(\hat{\boldsymbol{w}}_N^*)$) and performs poorer than the optimal portfolio from the noise space ($RV(\hat{\boldsymbol{w}}_N^*) \gg RV(\boldsymbol{w}_N^*)$).

Recall that the individual bottom eigenvectors might be poorly estimated, but the *span* of these eigenvectors is well estimated (i.e., space $\hat{\mathcal{N}}$ itself is well estimated). This reasoning is that $\hat{\mathcal{N}}$ is the space that is orthogonal to $\hat{\mathcal{S}}$, which is well estimated. Thus, we can expect good performance from a portfolio that depends only on space $\hat{\mathcal{N}}$ while being invariant to the precise configuration of the eigenvectors in $\hat{\mathcal{N}}$. The following example illustrates the case.

Example 2 (An extension of Example 1). From Equation (16), we know that the projected equal-weighted portfolio on the noise space weights the bottom two eigenvectors as follows:

$$\boldsymbol{w}_N^{EW} = \frac{\sum_{j=p-1}^{p} (\hat{\boldsymbol{v}}_j' \mathbf{1}) \hat{\boldsymbol{v}}_j}{\sum_{j=p-1}^{p} (\hat{\boldsymbol{v}}_j' \mathbf{1})^2}.$$

Note that this definition does not refer to eigenvalues at all. One property of this portfolio is that it is invariant to θ:

$$\begin{aligned}
\boldsymbol{w}_N^{EW} &= \frac{\hat{\boldsymbol{v}}_{p-1}' \mathbf{1}}{(\hat{\boldsymbol{v}}_{p-1}' \mathbf{1})^2 + (\hat{\boldsymbol{v}}_p' \mathbf{1})^2} \hat{\boldsymbol{v}}_{p-1} + \frac{\hat{\boldsymbol{v}}_p' \mathbf{1}}{(\hat{\boldsymbol{v}}_{p-1}' \mathbf{1})^2 + (\hat{\boldsymbol{v}}_p' \mathbf{1})^2} \hat{\boldsymbol{v}}_p \\
&= \frac{\rho(\cos\theta - \sin\theta)}{\rho^2(\cos\theta - \sin\theta)^2 + \rho^2(\cos\theta + \sin\theta)^2} (\boldsymbol{v}_{p-1} \cdot \cos\theta - \boldsymbol{v}_p \cdot \sin\theta) \\
&+ \frac{\rho(\cos\theta + \sin\theta)}{\rho^2(\cos\theta - \sin\theta)^2 + \rho^2(\cos\theta + \sin\theta)^2} (\boldsymbol{v}_{p-1} \cdot \sin\theta + \boldsymbol{v}_p \cdot \cos\theta) \\
&= \frac{1}{2\rho} \left(\boldsymbol{v}_{p-1} \cdot (\cos^2\theta + \sin^2\theta) + \boldsymbol{v}_p \cdot (\sin^2\theta + \cos^2\theta)\right) \\
&= \frac{1}{2\rho} (\boldsymbol{v}_{p-1} + \boldsymbol{v}_p) \\
&= \frac{\boldsymbol{v}_{p-1}' \mathbf{1}}{(\boldsymbol{v}_{p-1}' \mathbf{1})^2 + (\boldsymbol{v}_p' \mathbf{1})^2} \boldsymbol{v}_{p-1} + \frac{\boldsymbol{v}_p' \mathbf{1}}{(\boldsymbol{v}_{p-1}' \mathbf{1})^2 + (\boldsymbol{v}_p' \mathbf{1})^2} \boldsymbol{v}_p.
\end{aligned}$$

Using Equation (1), we find that:

$$RV(\boldsymbol{w}_N^{EW}) = \frac{1}{4\rho^2}\left(\xi + \frac{\xi}{\beta}\right) \approx RV(\boldsymbol{w}_N^*) \cdot \frac{\beta}{4} \quad \text{for } \beta \gg 1.$$

Again, the realized variance of $\boldsymbol{w}_N^{EW}$ is invariant to θ. We find that $\boldsymbol{w}_N^{EW}$ is comparable to the aggressive noise-only portfolio (Equation 3) in terms of realized variance, and that it can, in fact, be better than $\hat{\boldsymbol{w}}_N^*$ when $\frac{\pi}{4} \le \theta \le \frac{3\pi}{4}$. This finding makes sense because $\frac{\pi}{4} \le \theta \le \frac{3\pi}{4}$ means that $\hat{\boldsymbol{v}}_{p-1}$ is closer to $\boldsymbol{v}_p$ than $\boldsymbol{v}_{p-1}$.

Appendix B: Proofs

Proposition 3.1(Eigenvalue Concentration)

Proof. By Weyl's inequality, $|\lambda_i - \hat{\lambda}_i| \le \|\Sigma - \hat{\Sigma}\|_{op}$. Dividing both sides by λ_i proves the proposition. □

Lemma 3.3 (Portfolio Decomposition)

Proof. Using the Lagrangian multiplier method, we can easily find:

$$\boldsymbol{w}^* = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} = \frac{\sum_i \frac{\boldsymbol{v}_i'\mathbf{1}}{\lambda_i}\boldsymbol{v}_i}{\sum_i \frac{(\boldsymbol{v}_i'\mathbf{1})^2}{\lambda_i}},$$

where we use $\Sigma^{-1} = \sum_i (1/\lambda_i)\boldsymbol{v}_i\boldsymbol{v}_i'$. Similarly, we have:

$$\boldsymbol{w}_S^* = \frac{\sum_{j=1}^k \frac{\boldsymbol{v}_j'\mathbf{1}}{\lambda_j}\boldsymbol{v}_j}{\sum_{j=1}^k \frac{(\boldsymbol{v}_j'\mathbf{1})^2}{\lambda_j}}, \qquad RV(\boldsymbol{w}_S^*) = \frac{1}{\sum_{j=1}^k \frac{(\boldsymbol{v}_j'\mathbf{1})^2}{\lambda_j}}, \qquad \frac{1}{RV(\boldsymbol{w}_S^*)}\boldsymbol{w}_S^* = \sum_{j=1}^k \frac{\boldsymbol{v}_j'\mathbf{1}}{\lambda_j}. \tag{4}$$

Repeat this process for $\boldsymbol{w}_N^*$, and some algebraic manipulations yield Equation (2). □

Proposition 4.1(Bounding Realized Variance of any Portfolio from the Noise Space)

Proof. Because the noise space $\hat{\mathcal{N}}$ is spanned by $\hat{\boldsymbol{v}}_{k+1}, \ldots, \hat{\boldsymbol{v}}_p$, any vector $\boldsymbol{w}$ from $\hat{\mathcal{N}}$ can be presented as a linear combination of this basis, namely,

$$\boldsymbol{w} = (\hat{\boldsymbol{v}}_{k+1}, \ldots, \hat{\boldsymbol{v}}_p)\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = \hat{N}\boldsymbol{a}.$$

From the orthonormality of eigenvectors, we have:

$$||\boldsymbol{w}||_2^2 = \boldsymbol{w}'\boldsymbol{w} = \boldsymbol{a}'\hat{N}'\hat{N}\boldsymbol{a} = \boldsymbol{a}'\boldsymbol{a} = ||\boldsymbol{a}||_2^2. \tag{5}$$

Meanwhile, the definition of the noise bound, m, guarantees that the following inequality holds for any vector $\boldsymbol{b} \in \mathbb{R}^n$ such that $||\boldsymbol{b}||_2 = 1$,

$$\boldsymbol{b}'\left(\hat{N}'(\Sigma - \hat{\Sigma})\hat{N}\right)\boldsymbol{b} \le m. \tag{6}$$

Plugging $\boldsymbol{a}/||\boldsymbol{a}||_2$ into the previous inequality, we have

$$\left(\frac{\boldsymbol{a}}{||\boldsymbol{a}||_2}\right)'\left(\hat{N}'(\Sigma-\hat{\Sigma})\hat{N}\right)\left(\frac{\boldsymbol{a}}{||\boldsymbol{a}||_2}\right)\leq m.$$

Rearranging, we get:

$$(\hat{N}\boldsymbol{a})'\Sigma(\hat{N}\boldsymbol{a})\leq(\hat{N}\boldsymbol{a})'\hat{\Sigma}(\hat{N}\boldsymbol{a})+m||\boldsymbol{a}||_2^2.$$

Substituting $\boldsymbol{w}=\hat{N}\boldsymbol{a}$ and $||\boldsymbol{w}||_2^2=||\boldsymbol{a}||_2^2$ into the preceding inequality proves the proposition. $\square$

Lemma 6.1 (Projection Portfolios)

Proof. Clearly, $\boldsymbol{w}_S$ and $\boldsymbol{w}_N$ as defined in Equation (12) satisfy $\boldsymbol{w}_S\in\hat{\mathcal{S}}$, $\boldsymbol{w}_S'\boldsymbol{1}=1$ and $\boldsymbol{w}_N\in\hat{\mathcal{N}}$, $\boldsymbol{w}_N'\boldsymbol{1}=1$. Combining $\hat{S}\hat{S}'+\hat{N}\hat{N}'=I$ with $\boldsymbol{w}'\boldsymbol{1}=1$, we have

$$1=\boldsymbol{w}'\boldsymbol{1}=\boldsymbol{w}'(\hat{S}\hat{S}'+\hat{N}\hat{N}')\boldsymbol{1}=\theta+\boldsymbol{w}'\hat{N}\hat{N}'\boldsymbol{1},$$

which implies $1-\theta=\boldsymbol{w}'\hat{N}\hat{N}'\boldsymbol{1}$. Plugging this equation into the right-hand side of Equation (11),

$$RHS=\hat{S}\hat{S}'\boldsymbol{w}+\hat{N}\hat{N}'\boldsymbol{w}=\boldsymbol{w}=LHS.$$

In this way, we prove that Equation (12) gives one solution. Assume that there is another solution,

$$\boldsymbol{w}=\tilde{\theta}\tilde{\boldsymbol{w}}_S+(1-\tilde{\theta})\tilde{\boldsymbol{w}}_N.$$

Then, we have

$$\theta\boldsymbol{w}_S-\tilde{\theta}\tilde{\boldsymbol{w}}_S=-(1-\theta)\boldsymbol{w}_N+(1-\tilde{\theta})\tilde{\boldsymbol{w}}_N.$$

The left-hand side belongs to $\hat{\mathcal{S}}$ while the right-hand side belongs to $\hat{\mathcal{N}}$. Because $\hat{\mathcal{S}}\cap\hat{\mathcal{N}}=\boldsymbol{0}$, both sides are $\boldsymbol{0}$. However, $\boldsymbol{w}_S'\boldsymbol{1}=\tilde{\boldsymbol{w}}_S'\boldsymbol{1}=1$. Therefore, the following holds:

$$0=\boldsymbol{0}'\boldsymbol{1}=(\theta\boldsymbol{w}_S-\tilde{\theta}\tilde{\boldsymbol{w}}_S)'\boldsymbol{1}=\theta-\tilde{\theta}.$$

The equation implies that $\boldsymbol{w}_S=\tilde{\boldsymbol{w}}_S$ and $\boldsymbol{w}_N=\tilde{\boldsymbol{w}}_N$. $\square$

Lemma 6.2 (The Solution to the Robust Optimization)

Proof. Because $\boldsymbol{w}\in\hat{\mathcal{N}}$, we have $\boldsymbol{w}=\hat{N}\boldsymbol{a}$. Thus,

$$\max_{\Psi\in\mathcal{U}}\boldsymbol{w}'\Psi\boldsymbol{w}=\max_{\Psi\in\mathcal{U}}\boldsymbol{a}'\hat{N}'\Psi\hat{N}\boldsymbol{a}=b\boldsymbol{a}'I_{n-k+1}\boldsymbol{a}.$$

The last equality holds because of the definition of the uncertainty set. Then Equation (14) becomes

$$\begin{aligned}\min_{\boldsymbol{a}}\quad & b\boldsymbol{a}'I_{n-k+1}\boldsymbol{a},\\ \text{subject to}\quad & \boldsymbol{a}'(\hat{N}'\boldsymbol{1})=1.\end{aligned}$$

Its solution is

$$\boldsymbol{a}^* = \frac{\hat{N}'\mathbf{1}}{\mathbf{1}'\hat{N}\hat{N}'\mathbf{1}},$$

which implies that the solution to the robust optimization is

$$\hat{N}\boldsymbol{a}^* = \frac{\hat{N}\hat{N}'\mathbf{1}}{\mathbf{1}'\hat{N}\hat{N}'\mathbf{1}}.$$

From Equation (12), the projection portfolio of the equal-weighted portfolio on $\hat{\mathcal{N}}$ is:

$$\boldsymbol{w}_N^{EW} = \frac{\hat{N}\hat{N}'(\mathbf{1}/p)}{(\mathbf{1}/p)'\hat{N}\hat{N}'\mathbf{1}} = \frac{\hat{N}\hat{N}'\mathbf{1}}{\mathbf{1}'\hat{N}\hat{N}'\mathbf{1}} = \hat{N}\boldsymbol{a}^*. \quad \square$$