The Limits of Diversification

Ludger Hentschel

April 14, 2025

Abstract
Diversification offers clear benefits by reducing risk and improving Sharpe ratios. However, positive correlations among assets or signals limit these advantages. Correlations not only slow the pace at which diversification reduces risk but also impose a hard upper limit on its ultimate effectiveness. In the absence of correlations, diversification benefits grow without bound in proportion to the square root of the number of signals. This can be highly misleading when the number of signals is large, even if the correlations are small.

When all signals share a common correlation, the upper bound of the diversification benefits is equal to the inverse of the square root of the common correlation. For instance, with a low common correlation of 0.1, even an infinite number of signals roughly triple the Sharpe ratio of a single signal. When the common correlation rises to 0.8, diversification benefits are limited to a mere 12 percent.

We should take advantage of diversification benefits, especially because those benefits can often be obtained at relatively low costs. Moreover, numerically moderate multiples of Sharpe ratios can be economically very important. Still, it is unlikely that a portfolio of many weak signals can achieve a very high Sharpe ratio based on diversification alone. Achieving a high portfolio Sharpe ratio requires strong constituent signals. Since diversification has a multiplicative effect on Sharpe ratios, starting from a higher base amplifies its absolute benefits.

www.ludgerhentschel.com   ludger@ludgerhentschel.com

Contents

Acknowledgements

I am grateful for helpful comments from John Y. Campbell and Ravi Jagannathan.

1 Introduction

Diversification offers clear benefits by reducing risk and enhancing Sharpe ratios. However, positive correlations among assets or signals limit these advantages. Not only do such correlations slow the benefits of diversification, but they also impose an upper bound on its ultimate effectiveness. When the number of signals N is large, the commonly used "square root of N" heuristic – correct for uncorrelated signals – can be highly misleading, even if the common correlation among signals is very small.

Investors often combine assets into a portfolio with the aim of achieving a high Sharpe ratio. Quantitative asset managers often combine multiple signal portfolios into an overall forecast.[1] When this approach is combined with machine learning methods, the number of such signal portfolios can quickly grow from a dozen signals to hundreds or thousands. Some asset managers report that they combine millions of signals.

Common intuition, built on the case of zero correlation, correctly suggests that the benefits of diversification increase with the number of signals. A natural generalization to the case with correlations is that the benefits grow more slowly with the number of signals. Far less obvious is that pervasive positive correlations impose an upper limit on the benefits of diversification, making numerically large diversification benefits impossible.

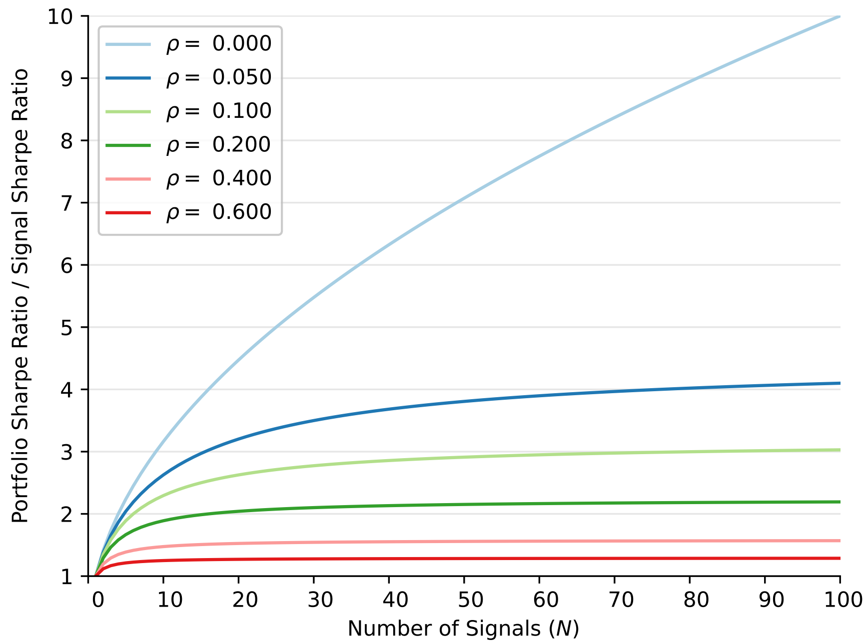
The benefits of diversification have a multiplier effect on Sharpe ratios. When all signals have the same common correlation with each other, diversification benefits have an upper bound equal to the inverse of the square root of the common correlation among the signals. For low common signal correlations of 0.1, an infinite number of signals cannot do much more than triple the Sharpe ratio of a single signal. When the common correlations are 0.8, the diversification benefits are capped at a 12 percent improvement.

As a result, it is unlikely that a portfolio of many weak signals can achieve a very high Sharpe ratio based on diversification. To achieve a high portfolio Sharpe ratio, at least some of the individual signals must have fairly high Sharpe ratios. An additional benefit is that applying the multiplier effect of diversification to signals with higher Sharpe ratios produces larger absolute increases in portfolio Sharpe ratios.

Figure 1 illustrates the ratio of maximum Sharpe ratios for a portfolio with N signals that have identical Sharpe ratios and a portfolio with 1 signal. If the common correlations ρ are zero, the top line in the figure shows that the Sharpe ratio continues to increase with the square root of the number

[1] The analysis applies equally to portfolios of assets or portfolios built on signal portfolios that form an overall active trading strategy. We use the terms assets and signals interchangeably.

Figure 1: Sharpe Ratio Improvements from Diversification



The figure shows the Sharpe ratio benefits from diversification across N signals with the same Sharpe ratio θ and a common correlation ρ with all other signals.

A portfolio of N signals achieves a multiple of the individual Sharpe ratios

$$\frac{\theta_p(N)}{\theta} = \sqrt{\frac{N}{1+\rho(N-1)}} \qquad \text{with} \quad \lim_{N\to\infty} \frac{\theta_p(N)}{\theta} = \frac{1}{\sqrt{\rho}}.$$

of signals. For pervasive positive correlations, however, the figure clearly shows that the Sharpe ratio improvement converges to an upper limit and that the upper limit falls as the common correlations among the signals rise. Even if the pervasive correlations are very small, the presence of correlations has a dramatic effect on the attainable benefits of diversification.

Even numerically moderate increases in the Sharpe ratio can be economically large. We should take advantage of the diversification benefits that are available. But we should be realistic about how large the benefits of diversification can be. Modern machine learning methods provide access to a very large number of signals at relatively low cost. But the diversification benefits of these signals are likely to be limited unless the signals become better, by incorporating new information, instead of merely presenting correlated variations of existing ideas.

Although there are limits on the benefits of diversification, approaching those limits almost always requires multiple signals. Even when signal correlations are 0.9, for example, simply picking one of the signals sacrifices diversification benefits. To obtain 90% of the available diversification benefits, we generally require more than 10 of the correlated signals, even if the common correlations are 0.9 or higher.

If some signals have negative correlations with others, this clearly increases the available diversification benefits and we should be alert to such exceptional cases. However, the absence of arbitrage requires that such negative correlations must be the exception, not the norm, when we have a large number of signals. As long as the typical correlation among signals is positive, our main qualitative results apply.

The remainder of this note provides the analysis and intuition behind these results. We first establish the main results using a simple correlation structure and then generalize the results to richer correlation structures.

2 Diversification in Optimal Portfolios

Under simplifying but natural assumptions, we can derive analytical results for the Sharpe ratio benefits of diversification. The analysis applies equally to portfolios of assets and combinations of signals. We interpret the constituents as signal portfolios and refer to them as signals for brevity. This situation arises when we forecast returns in several different ways and then combine these signal portfolios into an overall portfolio. Even though the signal portfolios may actively trade the underlying assets, it is sensible to consider the signal portfolios as static ingredients if their return and risk characteristics are stable.

These signal portfolios are familiar from many approaches to systematic investing. For example, value and momentum portfolios based on the work of Fama and French (1992) and Jegadeesh and Titman (1993) fit into this framework. The signals can be fundamental, like valuations, or purely technical, like momentum. Some of the signals can be passive, in the sense of an index weighted by market capitalization. The signals can be constructed by investment professionals or by machine learning algorithms, as those reviewed by Nissim (2022). Alternatively, we can interpret the signal portfolios as portfolios managed by investors and assembled into a portfolio by an allocator at a pension plan or a multi-manager fund.

2.1 The Signals

If we have a collection of N signals, we can combine these signals into an optimal portfolio with a higher Sharpe ratio. If the individual signals have expected excess returns μ_i and risk σ_i, their Sharpe ratios are $\theta_i = \mu_i/\sigma_i$.[2] In order to focus on the effects of correlation among these signals, we assume that all of the signals have the same expected excess return μ, risk σ, and hence Sharpe ratio θ. We maintain this important assumption throughout.

Allowing for different Sharpe ratios generally reduces the benefits of diversification. If some signals have higher Sharpe ratios, the optimal portfolio allocates more to these signals and less to others. This portfolio concentration reduces the benefits of diversification. The optimal portfolio accepts higher risk in exchange for even higher expected return.

We assume that the signal Sharpe ratios are known or at least estimated without bias. The limits of diversification we discuss here are not estimation problems. The limits are also entirely distinct from the important multiple testing problems that frequently arise when working with a large number of empirical signals. Harvey, Liu, and Zhu (2016) discuss these multiple testing issues.

2.2 A Simple Correlation Structure

A natural way to parameterize common correlation among the signals is to assume that they all have the same pairwise correlation $0 \leq \rho < 1$, so that the correlation matrix is

$$\begin{aligned} \boldsymbol{R} &= \begin{bmatrix} 1 & \rho & \rho & \dots & \rho \\ \rho & 1 & \rho & \dots & \rho \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ \rho & \rho & \dots & \rho & 1 \end{bmatrix} \\ &= (1-\rho)\boldsymbol{I} + \rho\boldsymbol{\iota}\boldsymbol{\iota}', \end{aligned} \tag{1}$$

where $\boldsymbol{I}$ is the $N \times N$ identity matrix and $\boldsymbol{\iota}$ is an N-element vector of ones.[3]

Under this structure, the covariance matrix for the N signal portfolio returns is

$$\begin{aligned} \boldsymbol{\Sigma} &= \sigma^2 \boldsymbol{R} \\ &= \sigma^2 \left[(1-\rho)\boldsymbol{I} + \rho\boldsymbol{\iota}\boldsymbol{\iota}'\right]. \end{aligned} \tag{2}$$

[2] If expected returns or volatilities vary over time, we interpret μ and σ as conditional expectations at a point in time. To keep the notation simple, we don't indicate this with a subscript t, for example.

[3] Bold symbols denote vectors or matrices.

The constraint that the common correlation ρ is nonnegative seems sensible for related signals. It is also necessary in order to ensure that we have a proper (positive definite) correlation matrix for any N. For large N, the correlation matrix in equation (1) is positive definite only if $\rho \geq 0$. For negative ρ, we would be able to construct portfolios with negative variance, something we generally rule out based on first principles. Hoping for a large correlation matrix with positive and negative correlations clustered around zero is equivalent to hoping for arbitrage.

The correlation structure is based on strong symmetry among the signals. Clearly, this is restrictive in general. However, if we have multiple signals we wish to combine into a single forecast, it seems natural that these signals would have material overlap with each other.[4] Moreover, it even seems desirable to have a collection of signals with this type of structure. If we have clusters of signals with higher correlations within the clusters, the diversification benefits are lower, as we will show in the next section.

Because the focus here is on using analytical results to develop intuition about the benefits of diversification, we work with correlation structures that permit analytical solutions. For practical optimizations using numerical inputs, we can use any correlation structure.

One way to generate this correlation pattern is with a single-factor structure for the signals: all signals have equal correlation with the common factor but have mutually uncorrelated deviations with equal residual variance. This single-factor assumption is common when analyzing diversification for stock portfolios. In that case, the market portfolio is the natural common factor. This has been investigated extensively with empirical analysis going back to Evans and Archer (1968) and Elton and Gruber (1977). Elton and Gruber (1977) also derive analytical results. Unfortunately, those results and their derivation are much more complex than the analysis here. This literature focuses on equally weighted portfolios for simplicity without showing that this is reasonable.

Signal portfolios frequently are market neutral. As a result, the market is not a common factor for such portfolios. Many signal portfolios are variations on a common objective, say returns following analyst revisions. For such portfolios, the common factor most likely stems from the common objective. Moreover, the better the signals become, the more closely they match the common objective, and the larger their correlation. Hoping for

[4]For stocks, Elton and Gruber (1973) show that this correlation structure provides better predictions than standard estimates of the sample covariance matrix. Ledoit and Wolff (2004) use this fact when shrinking the sample covariance estimate toward this equal correlation structure in order to obtain better covariance estimates.

zero correlations seems unrealistic. Yet, when forecast accuracy is low, as is commonly the case for return predictions, individual forecasts come with a lot of noise. That noise is not obviously correlated across signals and likely can be reduced via diversification. The next section discusses this case in more detail.

2.3 The Optimal Portfolio

Given a collection of signals, we can find the weights that maximize the Sharpe ratio in an unconstrained portfolio. Since the Sharpe ratio is invariant under leverage, we focus on the fully invested portfolio with weights summing to 1.

General Portfolios

For portfolio weights $\boldsymbol{w}$, the expected excess return on the portfolio is

$$\mu_p = \boldsymbol{w}'\boldsymbol{\mu}, \tag{3}$$

where $\boldsymbol{\mu}$ is a column vector of expected excess returns for each of the signals. The portfolio variance is

$$\sigma_p^2 = \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w}. \tag{4}$$

The portfolio weights that maximize the portfolio Sharpe ratio are

$$\begin{aligned} \boldsymbol{w} &= \left(\boldsymbol{i}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right)^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \\ &= \alpha\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}, \end{aligned} \tag{5}$$

where α is a scale factor that ensures that the weights sum to 1. This portfolio is commonly called the "tangency portfolio".[5] In general, some of these weights could be negative, depending on $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}$.

Finally, the portfolio Sharpe ratio is

$$\begin{aligned} \theta_p &= \frac{\mu_p}{\sigma_p} \\ &= \left(\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w}\right)^{-1/2}\boldsymbol{w}'\boldsymbol{\mu}. \end{aligned} \tag{6}$$

By construction, the portfolio attains the maximum Sharpe ratio at the optimal weights

$$\theta_p = \left(\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w}\right)^{-1/2}\boldsymbol{w}'\boldsymbol{\mu}$$

[5] See Ingersoll (1987) or Zivot (2021) for detailed derivations of this standard result.

$$= \sqrt{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}. \tag{7}$$

These results hold for any collection of signals with expected excess returns $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$.[6]

In fundamental factor models, we generally orthogonalize signal exposures in order isolate the contributions from different signals. Orthogonalizing signal exposures does not orthogonalize signal returns. Although this is not commonly done, we could orthogonalize the signal returns. Equation (7) implies that the optimal Sharpe ratio is the same for all rotations of the signal portfolios. Although we can rotate the signals to reduce or eliminate the return correlations, doing so changes the means and covariances in a manner that leaves the optimal Sharpe ratio unchanged. For example, if we rotate the signals by $\boldsymbol{\Sigma}^{-1/2}$, the means change to $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}$ and their covariance is equal to the identity matrix. However, the optimal Sharpe ratio $\theta_p = \sqrt{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1/2}\boldsymbol{I}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}} = \sqrt{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}$ remains unchanged.

Portfolios for Symmetric Signals

If, as we have assumed, all signals have the same expected excess return, the same variance, and the same common correlation, we can further simplify the expression for the portfolio Sharpe ratio. Substituting $\boldsymbol{\mu} = \mu\boldsymbol{\iota}$ and $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{R}$ into the portfolio Sharpe ratio from equation (7), we find

$$\begin{aligned} \theta_p &= \sqrt{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}} \\ &= \frac{\mu}{\sigma}\sqrt{\boldsymbol{\iota}'\boldsymbol{R}^{-1}\boldsymbol{\iota}} \end{aligned} \tag{8}$$

and

$$\frac{\theta_p}{\theta} = \sqrt{\boldsymbol{\iota}'\boldsymbol{R}^{-1}\boldsymbol{\iota}}. \tag{9}$$

The diversification benefits θ_p/θ depend on the sum of all entries in the inverse correlation matrix $\boldsymbol{R}^{-1}$. This is true for any correlation structure in $\boldsymbol{R}$, as long as all signals have equal expected return and equal variance.[7]

Only from this point forward does the analysis rely on the structure of the correlation matrix.

[6] The optimal portfolio has a nonnegative Sharpe ratio because θ_p is a quadratic form involving the inverse covariance matrix. If the covariance matrix is positive (semi) definite, so is its inverse. All quadratic forms in positive definite matrices are positive.

[7] Even if the signals have equal mean and variance, the portfolio weights may be unequal, depending on $\boldsymbol{R}$.

Given the structure of the correlation matrix, it's analytical inverse is

$$\begin{aligned} \boldsymbol{R}^{-1} &= \frac{1}{1-\rho}\,\boldsymbol{I} - \frac{\rho}{(1-\rho)^2}\,\frac{1}{1+\rho/(1-\rho)\;\boldsymbol{\iota}'\boldsymbol{\iota}}\,\boldsymbol{\iota}\boldsymbol{\iota}' \\ &= \frac{1}{1-\rho}\,\boldsymbol{I} - \frac{\rho}{(1-\rho)^2+\rho(1-\rho)N}\,\boldsymbol{\iota}\boldsymbol{\iota}'. \end{aligned} \tag{10}$$

This inverse is a standard result sometimes called the Sherman-Morrison-Woodbury formula. (See Sherman and Morrison (1950) and Bartlett (1951).) In this inverse, the coefficients on $\boldsymbol{I}$ and $\boldsymbol{\iota}\boldsymbol{\iota}'$ are both positive if $\rho > 0$. As a result, the diagonal entries are positive and the off-diagonal entries are negative if $\rho > 0$.

The key product we need for the portfolio Sharpe ratio sums all of the elements in the inverse correlation matrix,

$$\begin{aligned} \boldsymbol{\iota}'\boldsymbol{R}^{-1}\boldsymbol{\iota} &= \frac{1}{1-\rho}\,N - \frac{\rho}{(1-\rho)^2+\rho(1-\rho)N}\,N^2 \\ &= \frac{N}{1+\rho(N-1)}. \end{aligned} \tag{11}$$

In this sum, the positive contributions from the diagonal entries in the first term are partly offset by negative contributions from off-diagonal entries in the second term.

As the number of signals goes to infinity, the sum of the elements converges to

$$\lim_{N\to\infty} \frac{N}{1+\rho(N-1)} = \frac{1}{\rho}. \tag{12}$$

Such limits exist for all correlation matrices for which the sum of the elements in the inverse does not vanish as N grows. For positive correlations, the limit is finite, unless the correlation matrix becomes band diagonal as N grows, with a narrow band of non-zero correlations. (See appendix A for additional details.)

2.4 Comparing Portfolios to Signals

The algebra above implies that the maximum portfolio Sharpe ratio has an upper limit

$$\lim_{N\to\infty} \theta_p(N) = \frac{\mu}{\sigma}\frac{1}{\sqrt{\rho}}. \tag{13}$$

As a result, the limiting portfolio Sharpe ratio relative to a single signal Sharpe ratio is

$$\lim_{N \to \infty} \frac{\theta_p(N)}{\theta} = \frac{1}{\sqrt{\rho}}. \tag{14}$$

If the signals are correlated, there is an upper limit to the Sharpe ratio improvement that we cannot breach – regardless of how many signals we combine.

For example, if the signals have a common correlation of $\rho = 0.25$, even a portfolio with a very large number of such signals can only double the Sharpe ratio of a single signal. In practice, correlations among signals are likely to be higher, possibly much higher, than 0.25. Of course, that quickly reduces the Sharpe ratio improvements we can possibly obtain from diversification. If $\rho = 0.8$, the maximum possible improvement shrinks to roughly 12 percent.

Combining equation (8) and equation (11), the portfolio Sharpe ratio relative to the signal Sharpe ratio for any N is

$$\frac{\theta_p(N)}{\theta} = \sqrt{\frac{N}{1 + \rho(N-1)}}. \tag{15}$$

We have expressed the benefits of diversification as a ratio of Sharpe ratios. This is convenient analytically. It is important to note, however, that any multiple implies a larger absolute improvement in Sharpe ratio if we start from a higher base. In an absolute sense, the benefits of diversification are larger for signals with higher individual Sharpe ratios.

Figure 1 illustrates the diversification benefits in equation (15) for different values of ρ and N. The figure clearly shows the lower maximum diversification benefits achievable in the presence of higher correlations. The figure also shows that diversification benefits reach their upper limits at moderate numbers of signals. For higher correlations, we approach the upper limits of the diversification benefits quite quickly.

Zero Correlation

If all correlations are zero, the diversification benefit is

$$\left.\frac{\theta_p(N)}{\theta}\right|_{\rho=0} = \sqrt{N}. \tag{16}$$

This result commonly appears in discussions of diversification, possibly because it is easier to derive.[8] Unfortunately, this may give the impression that the benefits of diversification are limited only by the maximum available N. This is only true if $\rho = 0$. The top line in figure 1 is for $\rho = 0$. Comparing this line to the others makes clear how misleading intuition based on $\rho = 0$ can be, even if the common correlations are relatively small. In the presence of pervasive correlation, $\rho > 0$, there is a maximum diversification benefit we cannot exceed and it is generally far below $\sqrt{N}$.

Adequate Diversification

We can also derive the number of signals we should average to achieve a given percentage of the possible gains. For a fraction π of the possible gains, we can solve

$$\begin{aligned} \frac{\theta_p(N)}{\theta} - 1 &\geq \pi \left[\frac{\theta_p(\infty)}{\theta} - 1 \right] \\ \sqrt{\frac{N}{1+\rho(N-1)}} - 1 &\geq \pi \left[\frac{1}{\sqrt{\rho}} - 1 \right]. \end{aligned} \tag{17}$$

Unfortunately, the analytical solution for N does not provide easy intuition.[9] Instead of studying the analytical result, we can summarize the solution graphically.

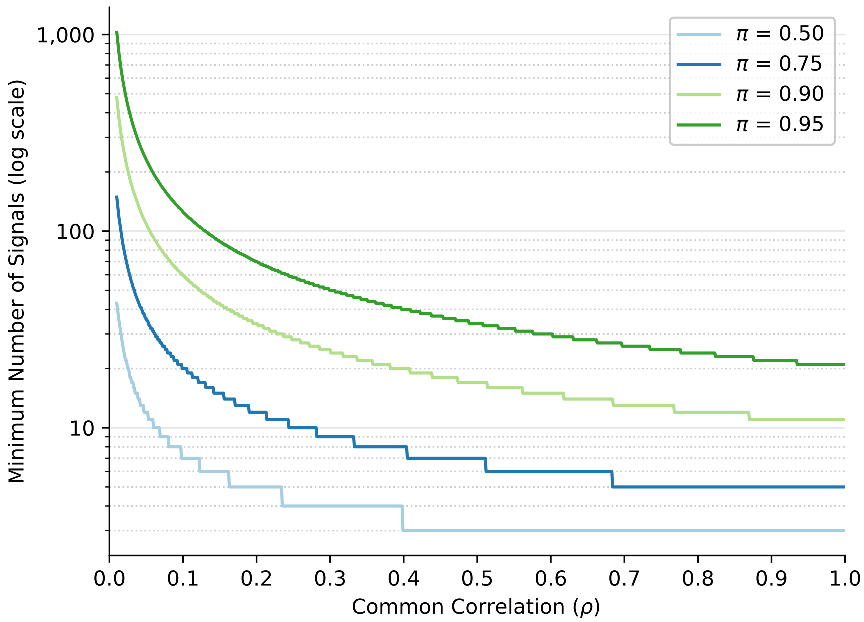
Figure 2 shows the number of signals that are sufficient to achieve 50, 75, 90, or 95 percent of the possible diversification benefits for different correlations ρ. The figure shows that the required number of signals falls as the common correlation across signals rises. For $\rho = 0.25$, we need 28 signals to achieve 90 percent of the potential benefits. For $\rho = 0.8$, we only need 12 signals to achieve 90 percent of the potential benefits from diversification. For plausible correlations, diversification across 10 to 100 signals appears to achieve the vast majority of the total available diversification benefits. Even with high correlations, $\rho > 0.9$, however, roughly 10 signals are required to achieve 90 percent of the maximum diversification benefit.

[8] In a widely cited result, Clarke, de Silva, and Thorley (2006) describe such a "root N" law. While the authors carefully point out that this result only holds for a diagonal correlation matrix, the common interpretation of this and similar results seems to omit this important caveat on the correlation structure.

[9] Ignoring integer constraints on N, the solution is

$$N \geq 1 + \frac{1}{\rho} \frac{\pi(2-\pi)\sqrt{\rho} + \pi^2}{(1-\pi)^2\sqrt{\rho} + (1-\pi^2)}.$$

Figure 2: Number of Signals Required for Adequate Diversification



The figure shows the number of signals required to achieve 50, 75, 90, and 95 percent of the maximum possible increase in Sharpe ratio due to diversification. All signals have the same Sharpe ratio and share a common correlation ρ.

The lines graph the smallest integer N that satisfies

$$\sqrt{\frac{N}{1+\rho(N-1)}} - 1 \geq \pi \left[\frac{1}{\sqrt{\rho}} - 1 \right], \qquad \pi \in \{0.50, 0.75, 0.90, 0.95\},$$

where the left side is the actual diversification benefit, for a given number of signals N, and the right side is π times the maximum possible diversification benefit.

Machine Learning Ensembles

In the machine learning context, a combination of forecasts like equation (5) is commonly called ensemble averaging, Bayesian model averaging, or a stacked model. See Wolpert (1992), Breiman (1996), Hoetig, Madigan, Raferty, and Volinsky (1999), or Hastie, Tibshirani, and Friedman (2008) for discussions.

Opitz and Maclin (1999) use 23 different data sets to empirically demonstrate that ensembles of machine learning predictions follow the pattern shown in figure 1.[10] They obtain this common empirical result because the correlations among the machine learning predictions in their ensembles

[10]In machine learning forecasting applications outside finance, it is common to track the reduction in the forecast error volatility. For unbiased predictors, the reduction in forecast error standard deviation is proportional to the inverse of the Sharpe ratio improvement shown in figure 1.

are mostly positive, even though isolated correlation estimates are slightly negative.

Magnus and Vasnev (2023) use a very similar framework to show that confidence intervals for forecast combinations computed based on the assumption that the forecasts are uncorrelated can be far too small. They focus on the effects of rising correlations among a small number of forecasts. Our focus is on the effects of low correlations among a large number of signals.

In general, we don't know the optimal combination of signals and must estimate the signal weights, possibly by estimating the covariance matrix $\boldsymbol{\Sigma}$. In this case, estimation error in the covariance matrix can introduce so much noise that we are unable to derive any diversification benefits. Applying shrinkage to the covariance estimate, as in Ledoit and Wolff (2004), or imposing positivity constraints on the weights, as in Breiman (1996), can be very helpful. Jagannathan and Ma (2003) show that these two approaches are closely related.

Here, we can derive the optimal signal weights based on known structures. If we have to estimate the signal weights, the available benefits of diversification are probably lower than the bounds presented here because we are unlikely to find the truly optimal signal blend.

Implied Correlations

Inverting equation (15) can yield interesting insight into the effective correlation among signals in a portfolio. Given estimates of the Sharpe ratio improvement and the number of signals, the implied correlation is

$$\widehat{\rho} = \frac{N\theta^2 - \theta_p^2(N)}{N\theta_p^2(N) - \theta_p^2(N)} \tag{18}$$

with a limit of

$$\lim_{N\to\infty} \widehat{\rho} = \frac{\theta^2}{\theta_p^2(N)}. \tag{19}$$

For example, if we estimate that the portfolio Sharpe ratio is twice that of a typical signal and we have 10 signals, then the implied correlation is 0.16. For a large number of signals, the implied correlations would equal one quarter.

2.5 Alternative Interpretation

Another way to understand these results is to realize that the optimal portfolio is equally weighted. This follows from the symmetry of the problem: We have assumed that all signals have the same expected excess return, the same variance, and the same common correlation.[11]

We can also show that the optimal weights for this case are equal by substituting $\boldsymbol{\mu} = \mu\boldsymbol{\iota}$ and $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{R}$ into the signal weights from equation (5),

$$\begin{aligned} \boldsymbol{w} &= \alpha\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \\ &= \alpha\left(\sigma^2\boldsymbol{R}\right)^{-1}\mu\boldsymbol{\iota} \\ &= \alpha\frac{\mu}{\sigma^2}\boldsymbol{R}^{-1}\boldsymbol{\iota} \\ &= \left[\frac{1}{1-\rho} - \frac{\rho\, N}{(1-\rho)^2 + \rho(1-\rho)N}\right]\boldsymbol{\iota}. \end{aligned} \tag{20}$$

The variance of the equally weighted portfolio is

$$\begin{aligned} \sigma_p^2(N) &= \frac{\sigma^2}{N^2}\,\boldsymbol{\iota}'\boldsymbol{R}\boldsymbol{\iota} \\ &= \frac{1}{N}\sigma^2 + \frac{N-1}{N}\rho\sigma^2, \end{aligned} \tag{21}$$

which is a weighted average of the variance and the covariances. In the limit, this becomes

$$\lim_{N\to\infty}\ \sigma_p^2(N) = \rho\sigma^2, \tag{22}$$

since the variance terms diversify away and the portfolio variance converges to the average covariance. The fact that the portfolio variance converges to a lower limit greater than zero is the key driver limiting the benefits of diversification.

When all assets have the same expected excess return, the portfolio has the same expected excess return, $\mu_p = \mu$. With this additional fact, we can find the limiting portfolio Sharpe ratio via substitution,

$$\begin{aligned} \lim_{N\to\infty}\ \theta_p(N) &= \frac{\mu_p}{\sigma_p} \\ &= \frac{\mu}{\sigma}\,\frac{1}{\sqrt{\rho}}, \end{aligned} \tag{23}$$

as in equation (13).

[11] Khang, Picca, Zhang, and Zhu (2023) show that equal allocations to signal portfolios can lead to attractive outcomes even when they are not strictly optimal.

3 Richer Correlation Structures

The previous section derived analytical results for a relatively simple but plausible correlation structure. This section generalizes the results to a richer correlation structure with clusters of signals that have higher correlation within clusters and lower correlation across clusters. While this correlation structure is probably more realistic, the analytical solutions are slightly more complex. Importantly, the main result remains: Pervasive correlation imposes an upper limit on the benefits of diversification.

3.1 Block-Diagonal Correlations

We first analyze a block-diagonal correlation matrix where we have material correlation among related signals within a theme but zero correlation across different signal themes. In the following section, we will add correlation across signal themes.

A block-diagonal correlation matrix with k blocks is

$$\boldsymbol{R}_B = \begin{bmatrix} \boldsymbol{R}_1 & 0 & 0 & \dots & 0 \\ 0 & \boldsymbol{R}_2 & 0 & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \dots & 0 & \boldsymbol{R}_k \end{bmatrix}, \tag{24}$$

where $\boldsymbol{R}_i$ is a correlation matrix for signals within theme i. For simplicity, we assume that all $\boldsymbol{R}_i$ have the same $n \times n$ size and have the same structure as equation (1). To unambiguously identify the correlations within clusters, we label them ϱ. If $\boldsymbol{R}_B$ has shape $N \times N$, then $N = kn$. Clearly, the block-diagonal correlation matrix has less pervasive correlation than $\boldsymbol{R}$ in equation (1).

For a block-diagonal matrix, the inverse is

$$\boldsymbol{R}_B^{-1} = \begin{bmatrix} \boldsymbol{R}_1^{-1} & 0 & 0 & \dots & 0 \\ 0 & \boldsymbol{R}_2^{-1} & 0 & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \dots & 0 & \boldsymbol{R}_k^{-1} \end{bmatrix}. \tag{25}$$

Applying the same calculations as in the previous section to each $\boldsymbol{R}_i^{-1}$, we find that

$$\iota' R_B^{-1} \iota = \frac{kn}{1 + \varrho(n-1)}. \tag{26}$$

Here too, we see that a positive definite correlation matrix in equation (24) implies constraints on ϱ. In this case, we require $\varrho > -1/(n-1)$. As n grows, this approaches $\varrho > 0$.

As the number of signals $N = kn$ goes to infinity, it matters whether k is rising faster than n or slower than n. If only n is increasing

$$\lim_{n\to\infty} \frac{kn}{1 + \varrho(n-1)} = \frac{k}{\varrho}. \tag{27}$$

In this case, the limiting portfolio Sharpe ratio relative to a single signal Sharpe ratio is

$$\lim_{n\to\infty} \frac{\theta_p(kn)}{\theta} = \frac{\sqrt{k}}{\sqrt{\varrho}}. \tag{28}$$

Despite the much more limited correlation, there remains an upper limit of diversification if we don't increase the number of uncorrelated signal themes. However, due to the less pervasive correlation, the upper limit is higher by a factor of $\sqrt{k}$ compared to equation (13) if $\varrho = \rho$. Of course, we generally believe that correlations within clusters are higher than pervasive correlations among all signals.

Due to the zero correlation across clusters, this structure allows unlimited gains from diversification if we can continue to add new clusters of n uncorrelated signals with the same Sharpe ratio. As k grows, the correlation matrix R_B and its inverse R_B^{-1} become dominated by zero correlations across themes. Clearly, finding new signal themes is harder than finding variations on existing signals. However, the portfolio benefits are much larger.

3.2 Correlated Clusters

A more realistic case has fairly high correlations within the clusters and relatively low but positive correlations across the clusters. We can solve this case by combining the analytical methods from the two cases we have analyzed previously.

A block-diagonal correlation matrix with correlation across blocks is

$$\boldsymbol{R}_C = \begin{bmatrix} \boldsymbol{R}_1 & \boldsymbol{\rho} & \boldsymbol{\rho} & \dots & \boldsymbol{\rho} \\ \boldsymbol{\rho} & \boldsymbol{R}_2 & \boldsymbol{\rho} & \dots & \boldsymbol{\rho} \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ \boldsymbol{\rho} & \boldsymbol{\rho} & \dots & \boldsymbol{\rho} & \boldsymbol{R}_k \end{bmatrix} \tag{29}$$

with

$$\boldsymbol{\rho} = \rho \boldsymbol{\iota}_n \boldsymbol{\iota}_n', \tag{30}$$

where $\boldsymbol{\iota}_n$ is an n-dimensional vector of ones and $\boldsymbol{R}_i$ is a correlation matrix for signals within theme i, as before. This correlation matrix has correlations ϱ within the clusters along the diagonal and correlations ρ for all signals that are not in the same cluster. We generally think of $\rho < \varrho$.

It is helpful to write this correlation matrix using Kronecker products as

$$\begin{aligned} \boldsymbol{R}_C &= \boldsymbol{I}_k \otimes \widetilde{\boldsymbol{R}}_i + \rho \boldsymbol{\iota}_N \boldsymbol{\iota}_N' \\ &= \boldsymbol{A} + \rho \boldsymbol{\iota}_N \boldsymbol{\iota}_N' && (31) \\ \widetilde{\boldsymbol{R}}_i &= (1-\varrho)\boldsymbol{I}_n + (\varrho - \rho)\boldsymbol{\iota}_n \boldsymbol{\iota}_n'. && (32) \end{aligned}$$

Since these expressions use identity matrices and vectors of ones with different dimensions, we use subscripts to indicate the dimensions: $\boldsymbol{I}_k$ is a k-dimensional identity matrix and $\boldsymbol{\iota}_N$ is an N-element column vector of ones.

We can compute the inverse of the correlation matrix $\boldsymbol{R}_C$ by applying the Sherman-Morrison-Woodbury formula twice:

$$\boldsymbol{R}_C^{-1} = \boldsymbol{A}^{-1} - \frac{1}{1 + \rho \boldsymbol{\iota}_N' \boldsymbol{A}^{-1} \boldsymbol{\iota}_N} \rho \boldsymbol{A}^{-1} \boldsymbol{\iota}_N \boldsymbol{\iota}_N' \boldsymbol{A}^{-1} \tag{33}$$

$$\boldsymbol{A}^{-1} = \boldsymbol{I}_k \otimes \widetilde{\boldsymbol{R}}_i^{-1} \tag{34}$$

$$\widetilde{\boldsymbol{R}}_i^{-1} = \frac{1}{1-\varrho+\rho} \boldsymbol{I}_n - \frac{\varrho - \rho}{(1-\varrho+\rho)^2 + \rho(1-\varrho+\rho)n} \boldsymbol{\iota}_n \boldsymbol{\iota}_n'. \tag{35}$$

As before, we want the sum of the elements in the inverse correlation matrix

$$\boldsymbol{\iota}_N' \boldsymbol{R}_C^{-1} \boldsymbol{\iota}_N = \boldsymbol{\iota}_N' \boldsymbol{A}^{-1} \boldsymbol{\iota}_N - \frac{1}{1 + \rho \boldsymbol{\iota}_N' \boldsymbol{A}^{-1} \boldsymbol{\iota}_N} \rho \boldsymbol{\iota}_N' \boldsymbol{A}^{-1} \boldsymbol{\iota}_N \; \boldsymbol{\iota}_N' \boldsymbol{A}^{-1} \boldsymbol{\iota}_N$$

$$= s - \rho \frac{s^2}{1 + \rho s}$$
$$= \frac{s}{1 + \rho s}, \tag{36}$$

where

$$s = \boldsymbol{\iota}_N' \boldsymbol{A}^{-1} \boldsymbol{\iota}_N. \tag{37}$$

Recognizing the k-fold repetition of elements in $\boldsymbol{A}^{-1}$, the sum of the elements in $\boldsymbol{A}^{-1}$ is

$$s = \boldsymbol{\iota}_N' \left(\boldsymbol{I}_k \otimes \widetilde{\boldsymbol{R}}_i^{-1} \right) \boldsymbol{\iota}_N$$
$$= k \boldsymbol{\iota}_n' \widetilde{\boldsymbol{R}}_i^{-1} \boldsymbol{\iota}_n$$
$$= k \left[\frac{n}{1 + (\varrho - \rho)(n-1)} \right]. \tag{38}$$

Finally, we can find

$$\boldsymbol{\iota}_N' \boldsymbol{R}_C^{-1} \boldsymbol{\iota}_N = \frac{s}{1 + \rho s}$$
$$= \frac{kn}{1 + (\varrho - \rho)(n-1) + \rho k n} \tag{39}$$

and the ratio of Sharpe ratios

$$\frac{\theta_p(kn)}{\theta} = \sqrt{\boldsymbol{\iota}_N' \boldsymbol{R}_C^{-1} \boldsymbol{\iota}_N}$$
$$= \sqrt{\frac{kn}{1 + (\varrho - \rho)(n-1) + \rho k n}}, \tag{40}$$

similar to equation (9).[12]

We started with the assumption that the correlations within clusters, ϱ, are higher than the correlations across clusters, ρ. Under this assumption, the diversification benefits for $\boldsymbol{R}_C$ are lower than the results for $\boldsymbol{R}$ in the previous section and those for the purely block-diagonal case $\boldsymbol{R}_B$. There is higher correlation within the clusters than in $\boldsymbol{R}$. There is higher correlation across the clusters than in the strictly block-diagonal case $\boldsymbol{R}_B$. Higher correlations lower diversification benefits.

[12]Once more, the requirement of a positive definite correlation matrix in equation (29) imposes constraints on ϱ and ρ. We meet those constraints if $0 < \rho \leq \varrho < 1$, although this is not strictly required for a positive definite correlation matrix $\boldsymbol{R}_C$.

As the number of signals within each cluster increases, we once again reach a diversification limit

$$\lim_{n\to\infty} \frac{\theta_p(kn)}{\theta} = \sqrt{\frac{k}{\varrho + (k-1)\rho}}. \quad (41)$$

The more pervasive correlation reduces the benefits of diversification. This limit is strictly lower than equation (28) for the purely block-diagonal correlation matrix, if $k > 1$ and $\rho > 0$.

The purely block-diagonal limit in equation (28) is higher by a factor of $1 + (k-1)\rho/\varrho$. For a large number of clusters k, the correlation across clusters ρ imposes a large penalty on the possible diversification benefits from $\boldsymbol{R}_C$ compared to $\boldsymbol{R}_B$.

This diversification limit further simplifies to equation (28) if $\rho = 0$,

$$\lim_{n\to\infty} \left.\frac{\theta_p(kn)}{\theta}\right|_{\rho=0} = \frac{\sqrt{k}}{\sqrt{\varrho}}. \quad (42)$$

But, as before, the assumption of zero correlations across clusters seems unrealistically optimistic.

As the number of clusters increases, the correlation across clusters dominates and we return to the diversification limit in equation (13) from the previous section

$$\lim_{k\to\infty} \frac{\theta_p(kn)}{\theta} = \frac{1}{\sqrt{\rho}}. \quad (43)$$

Even if we believe that the correlation across signal clusters, ρ, is relatively low, there remains a hard upper limit on the benefits of diversification.

The diversification benefits now depend on four parameters: k, n, ϱ, and ρ. Instead of a large number of graphs, we produce a table with some illustrative values. Table 1 illustrates the benefits of diversification from equation (40) for several different parameter combinations. For comparison, the left column under the heading $\sqrt{kn}$ shows diversification benefits when the correlations across signals are zero, $\varrho = \rho = 0$.

The table intentionally shows some small values for ρ, the correlation across clusters. A large investable universe offers many degrees of freedom, so that it is theoretically possible to construct signal portfolios with low return correlations. Whether this is practical obviously depends on the particulars of the signals. Low correlations across clusters likely require careful searching for thematically different signal ideas.

Table 1: Benefits of Diversification with Correlated Clusters

n	$\sqrt{kn}$	$\varrho = 0.60$ $\rho: 0.01$	0.10	0.25	$\varrho = 0.75$ $\rho: 0.01$	0.10	0.25	$\varrho = 0.90$ $\rho: 0.01$	0.10	0.25
Panel A: k = 5										
1	2.24	2.18	1.83	1.49	2.18	1.83	1.49	2.18	1.83	1.49
5	5.00	2.63	2.13	1.70	2.44	2.02	1.64	2.28	1.93	1.59
10	7.07	2.71	2.18	1.73	2.48	2.05	1.67	2.29	1.95	1.61
20	10.00	2.75	2.21	1.75	2.50	2.07	1.68	2.30	1.95	1.61
Panel B: k = 10										
1	3.16	3.02	2.24	1.69	3.02	2.24	1.69	3.02	2.24	1.69
5	7.07	3.60	2.50	1.83	3.35	2.41	1.80	3.14	2.33	1.76
10	10.00	3.70	2.54	1.85	3.40	2.44	1.81	3.16	2.34	1.77
20	14.14	3.75	2.56	1.86	3.42	2.45	1.82	3.17	2.35	1.78
Panel C: k = 20										
1	4.47	4.08	2.58	1.83	4.08	2.58	1.83	4.08	2.58	1.83
5	10.00	4.79	2.77	1.91	4.49	2.71	1.89	4.24	2.65	1.87
10	14.14	4.91	2.80	1.92	4.55	2.73	1.90	4.26	2.66	1.88
20	20.00	4.97	2.81	1.93	4.58	2.74	1.90	4.27	2.67	1.88
Panel D: k = 100										
1	10.00	7.07	3.02	1.96	7.07	3.02	1.96	7.07	3.02	1.96
5	22.36	7.73	3.07	1.98	7.47	3.05	1.98	7.23	3.04	1.97
10	31.62	7.83	3.08	1.98	7.52	3.06	1.98	7.25	3.04	1.97
20	44.72	7.88	3.08	1.98	7.55	3.06	1.98	7.26	3.04	1.97
Panel E: k = 1,000										
1	31.62	9.53	3.15	2.00	9.53	3.15	2.00	9.53	3.15	2.00
5	70.71	9.68	3.15	2.00	9.63	3.15	2.00	9.57	3.15	2.00
10	100.00	9.70	3.15	2.00	9.64	3.15	2.00	9.58	3.15	2.00
20	141.42	9.71	3.15	2.00	9.64	3.15	2.00	9.58	3.15	2.00

The table shows the benefits from diversification for different levels of correlation and different numbers of signals.

All signals have the same expected return and the same variance. The correlation matrix $\boldsymbol{R}_C$ has k clusters with correlations ρ across signals in different clusters. Each cluster has n signals and correlations ϱ across signals within the same cluster.

The diversification benefit is the Sharpe ratio for the optimal portfolio of all signals divided by the Sharpe ratio of a single signal,

$$\frac{\theta_p(kn)}{\theta} = \sqrt{\frac{kn}{1 + (\varrho - \rho)(n-1) + \rho kn}}.$$

Table 1 confirms that, when correlations within clusters are higher than correlations across clusters, it is more beneficial to add more clusters than to add more signals within clusters. However, the incremental benefit of going beyond far beyond 100 signal clusters may be marginal, unless the average correlations across clusters are close to 0.

Similarly, the table shows that the incremental benefit of going from 10 to 20 signals within each cluster is marginal when signal correlations within clusters are material. We are approaching the ultimate limits of the available diversification benefits in the table. Table 1 suggests that the benefits of going beyond a few thousand signals are likely to be small, unless the additional signals have higher Sharpe ratios. For higher diversification benefits, we would have to have lower correlations – while maintaining the Sharpe ratios of the underlying signals.

Although the table shows that a tenfold increase in Sharpe ratios is possible, even if the signals are correlated, this seems challenging in practice. Table 1 shows that diversification benefits of this size require a large number of signal clusters with very low correlations across clusters. Experience suggests that the average correlation across clusters rises as the number of clusters increases. In practice, a doubling or tripling of Sharpe ratios via diversification should be possible but likely already requires thoughtful construction of signals with low correlations.

Table 1 also shows that realistic diversification benefits are far from the common $\sqrt{N}$ baseline assumption. In panel E, there are 1,000 clusters. Even ignoring the signals inside the clusters, $\sqrt{1,000} = 31.62$ materially overstates the benefits of diversification, even if $\rho = 0.01$. Counting the signals inside the clusters exacerbates this problem since $\sqrt{20,000} = 141.42$. For large N, diversification benefits estimated as $\sqrt{N}$ are often orders of magnitude too big.

The more general correlation structure in $\boldsymbol{R}_C$ emphasizes the main point: pervasive correlation imposes an upper limit on the ultimate benefits of diversification. The exact nature of the correlation matters for the level of the limit but not for its existence. Moreover, the illustrations suggest that, for plausible levels of correlations, the maximum diversification benefits are numerically modest and are reached for moderate numbers of clusters k and moderate numbers of signals within each cluster n.

There are clear benefits from diversification. Those benefits are often economically large but numerically moderate. Importantly, the benefits are limited for most realistic correlation structures.

3.3 Level of Correlations

Our analysis demonstrates that assuming zero correlations across a large number of signals implies vastly larger diversification benefits than are actually available when there are even small positive, pervasive correlations across signals. There is a natural question whether average signal correlations may be zero or even negative.

When the number of signals is small, it is certainly possible that the average correlation across signal returns is zero or negative. A prominent example is the value and momentum pair among equity signals. Over long periods, the returns associated with value and momentum signals have negative correlation.

As the number of signals rises, it seems more likely that the signal returns have positive correlations. Since the number of underlying assets is generally large, and the predictive power of the signals often limited, there should be space for many different signals to predict returns. However, that space shrinks and gets crowded when the number of signals exceeds the number of assets and when the signals all focus on a subset of returns: the predictable component of the underlying asset returns. This crowding among the signals must raise the average correlation among the signals.

Due to market efficiency, as defined by Fama (1970), a large part of the total asset return variation should be unpredictable. Good signals exclude these unpredictable return components. Although return signals generally have low correlations with returns, they likely have higher correlation with a small subset of returns, the predictable component. As a consequence, signals likely have higher correlations with each other than one might guess based on the low correlations with returns.

Until recently, the number of signals employed in empirical asset pricing studies or portfolio management was much smaller than the number of assets in the investable universe. This surely allowed for low correlations among the signals, although the actual correlation would have depended on the chosen signal mix. More recently, modern computing and machine learning methods have led to a much larger number of signals, often exceeding the number of assets in the investable universe. With that increase, the signals must pack more densely into the same space, likely leading to an increase in average correlations.

Unfortunately, it is impossible to make a general statement about the value of the average correlation across signal returns. Nonetheless, there are strong reasons to believe that the average correlation becomes positive as the number of signals grows very large. This is not an argument against using a large number of signals. However, it raises the importance of recognizing that, especially under these circumstances, diversification benefits are much more limited than under the zero correlation heuristic.

4 Summary

We derive the multiplicative Sharpe ratio benefits of diversification across assets or signals with pervasive correlation. The analysis shows that the benefits of diversification face strict limits, unless the typical correlations across signals are zero.

To focus on the effects of correlation, we assume that all signals have identical expected returns and variances. This is the best-case scenario for diversification. Differences in expected returns naturally reduce optimal diversification and its benefits.

If all signals have common correlations with each other, the maximum benefit from diversification is equal to 1 over the square root of the common correlation. At low correlation values, like $\rho = 0.1$, this says that we can increase Sharpe ratios by a factor of just over 3 via diversification. For $\rho = 0.8$, however, the maximum possible improvement shrinks to roughly 12 percent. In case signals can be grouped into clusters with higher correlations within clusters and lower correlations across clusters, we can interpret ρ as the lower correlation across clusters and the benefits of diversification are more limited.

If we wish to achieve high portfolio Sharpe ratios, we cannot realistically hope to do so via diversification across a large number of correlated signals with low Sharpe ratios. This may be especially pertinent when machine learning methods can be applied to generate a large number of signals at relatively low cost.

The most plausible way of attaining a high portfolio Sharpe ratio is to assemble signals that individually have fairly high Sharpe ratios. An additional advantage of the better starting point is that the multiplicative benefits of diversification deliver larger absolute improvements in the portfolio Sharpe ratio. Productive machine learning applications to investment signals should focus on improving the signals, not just on generating many correlated signals with similar Sharpe ratios.

These results do not detract from the fact that diversification often offers relatively cheap benefits. We should take advantage of these benefits. But we should not hope for unrealistically high Sharpe ratio improvements from diversification across a large number of weak signals.

5 References

Bartlett, Maurice S., 1951, An inverse matrix adjustment arising in discriminant analysis, *Annals of Mathematical Statistics* 22, 107–111.

Breiman, Leo, 1996, Stacked regressions, *Machine Learning* 24, 49–64.

Clarke, Roger, Harindra de Silva, and Steven Thorley, 2006, The fundamental law of active management, *Journal of Investment Management* 4, 54–72.

Elton, Edward J., and Martin J. Gruber, 1977, Risk reduction and portfolio size: An analytical solution, *Journal of Business* 50, 415–437.

Elton, Edwin J., and Martin J. Gruber, 1973, Estimating the dependence structure of share prices – implications for portfolio selection, *Journal of Finance* 28, 1203–1232.

Evans, John L., and Stephen H. Archer, 1968, Diversification and the reduction of dispersion: An empirical analysis, *Journal of Finance* 23, 761–767.

Fama, Eugene F., 1970, Efficient capital markets: A review of theory and empirical work, *Journal of Finance* 25, 383–417.

Fama, Eugene F., and Kenneth R. French, 1992, The cross-section of expected stock returns, *Journal of Finance* 47, 427–465.

Harvey, Campbell R., Yan Liu, and Heqing Zhu, 2016, ... and the cross-section of expected returns, *Review of Financial Studies* 29, 5–68.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman, 2008, *The Elements of Statistical Learning*, second edition (Springer, New York, NY).

Hoetig, Jennifer A., David Madigan, Adrian E. Raferty, and Chris T. Volinsky, 1999, Bayesian model averaging: A tutorial, *Statistical Science* 14, 382–417.

Ingersoll, Jonathan E., 1987, *Theory of Financial Decision Making* (Rowman & Littlefield, Lanham, MD).

Jagannathan, Ravi, and Tongshu Ma, 2003, Risk reduction in large portfolios: Why imposing the wrong constraints helps, *Journal of Finance* 58, 1651–1683.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 65–91.

Khang, Kevin, Antonio Picca, Shaojun Zhang, and Minzhi Zhu, 2023, How inefficient is the $1/N$ strategy for a factor investor?, *Journal of Investment Management* 21, 103–119.

Ledoit, Olivier, and Michael Wolff, 2004, Honey, I shrunk the sample covariance matrix, *Journal of Portfolio Management* 110–119.

Magnus, Jan R., and Andrey L. Vasnev, 2023, On the uncertainty of a combined forecast: The critical role of correlations, *International Journal of Forecasting* 39, 1895–1908.

Nissim, Doron, 2022, Big data, accounting information, and valuation, *Journal of Finance and Data Science* 8, 69–85.

Opitz, David, and Richard Maclin, 1999, Popular ensemble methods: An empirical study, *Journal of Artificial Intelligence Research* 11, 169–198.

Parzen, Emanuel, 1961, An approach to time series analysis, *Annals of Mathematical Statistics* 32, 951–989.

Sherman, Jack, and Winifred J. Morrison, 1950, Adjustment of an inverse matrix corresponding to a change in one element of a given matrix, *Annals of Mathematical Statistics* 21, 124–127.

Siddiqui, Mohammed M., 1958, On the inverse of the sample covariance matrix in a stationary autoregressive process, *Annals of Mathematical Statistics* 29, 585–588.

Verbyla, Arunas P., 1985, A note on the inverse covariance matrix of the autoregressive process, *Australian Journal of Statistics* 27, 221–224.

Wolpert, David H., 1992, Stacked generalization, *Neural Networks* 5, 241–259.

Zivot, Eric, 2021, Introduction to computational finance and financial econometrics with R, Working paper, University of Washington, Seattle, WA.

A Alternative Covariance Structures

The main text discusses two plausible cases of pervasive correlation. One in which all signals have equal correlations with each other. The other in which signals have high correlations within clusters and lower correlations across clusters. This appendix explores a correlation matrix with exponentially declining correlations for which we can derive analytical results.

Once again, the main result is that widespread correlation among portfolio constituents limits the ultimate benefits of diversification. While the exact limit depends on the correlation structure, a limit exist for all correlation structures in which the fraction of positive entries in the correlation matrix and its inverse does not go to zero as the number of elements rises.

A.1 Declining Correlations

A correlation matrix with declining correlations may be

$$\boldsymbol{R}_D = \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{N-1} \\ \rho & 1 & \rho & \dots & \rho^{N-2} \\ \rho^2 & \rho & \ddots & & \vdots \\ \vdots & & & \ddots & \rho \\ \rho^{N-1} & \rho^{N-2} & \dots & \rho & 1 \end{bmatrix}. \tag{44}$$

Such a structure implies that the signals have been sorted so that the signals' correlations decline with the distance between the signals in the list.

This correlation structure arises for time-series observations that follow a first-order autoregression. Siddiqui (1958), Parzen (1961), and Verbyla (1985) show that the inverse of this correlation matrix is tridiagonal with

$$\boldsymbol{R}_D^{-1} = \frac{1}{1-\rho^2} \begin{bmatrix} 1 & -\rho & 0 & \dots & 0 \\ -\rho & 1+\rho^2 & -\rho & \dots & 0 \\ 0 & -\rho & \ddots & & \vdots \\ \vdots & & & 1+\rho^2 & -\rho \\ 0 & 0 & \dots & -\rho & 1 \end{bmatrix}. \tag{45}$$

As before, we can compute

$$\boldsymbol{\iota}'\boldsymbol{R}_D^{-1}\boldsymbol{\iota} = 2\,\frac{\rho}{1+\rho} + N\,\frac{1-\rho}{1+\rho}, \tag{46}$$

so that

$$\frac{\theta_p(N)}{\theta} = \sqrt{2\,\frac{\rho}{1+\rho} + N\,\frac{1-\rho}{1+\rho}}. \tag{47}$$

As the number of signals, N, goes to infinity, the ratio of Sharpe ratios also goes to infinity. This happens because most of the correlations in $\boldsymbol{R}_D$ converge to zero as N goes to infinity. If we were to encounter this situation, we would enjoy the boundless benefits of diversification. Unfortunately, it is implausible that the average correlation across signals would continue to decline as we add more and more signals.

If the declining correlation pattern is appealing, we can specify that the lowest correlations are constant, regardless of N,

$$\rho^{N-1} = \underline{\rho}. \tag{48}$$

That also implies that the largest correlations rise with N,

$$\rho = \underline{\rho}^{1/(N-1)}. \tag{49}$$

This seems more realistic: As we add more signals, the lowest correlations don't change and the highest correlations rise. This should happen if we "fill in" more signals into an existing set of signals.

With this change,

$$\frac{\theta_p(N)}{\theta} = \sqrt{2\,\frac{\underline{\rho}^{1/(N-1)}}{1+\underline{\rho}^{1/(N-1)}} + N\,\frac{1-\underline{\rho}^{1/(N-1)}}{1+\underline{\rho}^{1/(N-1)}}}, \tag{50}$$

which converges to

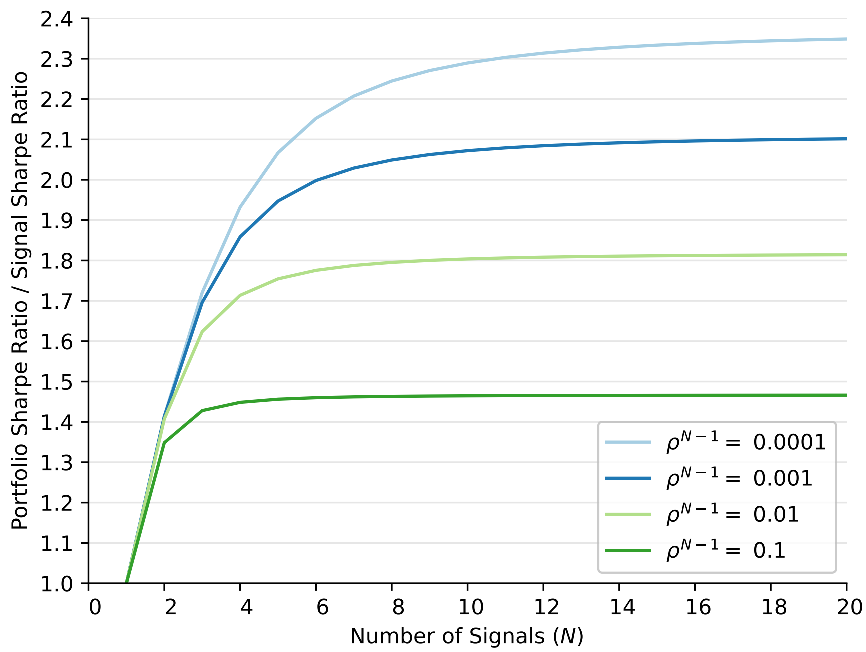
$$\lim_{N\to\infty} \frac{\theta_p(N)}{\theta} = \sqrt{1 - \frac{\ln(\underline{\rho})}{2}}. \tag{51}$$

Because $0 \le \underline{\rho} \le 1$, this limit is greater than 1. Certainly if $\underline{\rho}$ is greater than the common correlation in $\boldsymbol{R}$, the average correlation is higher and the limiting diversification benefit is lower than in equation (14).[13]

[13]Comparing the limits, we find that $\boldsymbol{R}_D$ provides lower ultimate diversification benefits than $\boldsymbol{R}$ when

$$\sqrt{1 - \frac{\ln(\underline{\rho})}{2}} < \frac{1}{\sqrt{\rho}}$$
$$\underline{\rho} > \exp\left\{2(\rho-1)/\rho\right\}.$$

Figure 3: Sharpe Ratio Improvements from Diversification



The figure shows the Sharpe ratio benefits from diversification across N signals with exponentially declining correlations. All of the individual signals have the same Sharpe ratio θ. The correlation between signals i and j is $R_D(i,j) = \rho^{|i-j|}$, as shown in equation (44). A portfolio of N signals achieves a multiple of the individual Sharpe ratios

$$\frac{\theta_p(N)}{\theta} = \sqrt{2\,\frac{\rho}{1+\rho} + N\,\frac{1-\rho}{1+\rho}}.$$

If we hold $\rho^{N-1} = \underline{\rho}$ constant, this ratio approaches

$$\lim_{N\to\infty} \frac{\theta_p(N)}{\theta} = \sqrt{1 - \frac{\ln(\underline{\rho})}{2}}.$$

Figure 3 illustrates the gains from diversification for this case. In the figure, some of the correlations decline to very low values. However, the average correlations are fairly high and thereby limit the gains from diversification even more than in figure 1.

When $\underline{\rho}$ is very small, visual inspection of the correlation matrix $\boldsymbol{R}_D$ might suggest that the correlation matrix has a banded diagonal structure. There are positive correlations close to the main diagonal but they fade to zero in the upper and lower triangles. Despite this appearance, there can be enough correlation among the signals to impose an upper limit on the benefits of diversification.

For plausible values of ρ, this is true even when $\underline{\rho}$ is very small.