Honey, I Shrunk the Sample Covariance Matrix

Problems in mean-variance optimization.

Olivier Ledoit and Michael Wolf

Since the seminal work of Markowitz [1952], mean-variance optimization has been the most rigorous way to pick stocks. The two fundamental ingredients are the expected (excess) return for each stock, which represents the portfolio manager's ability to forecast future price movements, and the covariance matrix of stock returns, which represents risk control.

To further specify the problem, in the real world most asset managers are not allowed to sell short, and in the modern world they are typically measured against the benchmark of an equity market index with fixed (or infrequently rebalanced) weights. There is fast and accurate quadratic optimization software that can solve this problem—provided it is fed the right inputs.

Estimating the covariance matrix of stock returns has always been one of the stickiest points. The standard statistical approach is to gather a history of past stock returns and compute their *sample covariance matrix*. Unfortunately, this creates problems that are well documented (see Jobson and Korkie [1980]).

To put it as simply as possible, when there are many stocks under consideration, especially compared to the number of historical return observations available (as is the usual case), the sample covariance matrix is estimated with a lot of error. This means the most extreme coefficients in such an estimated matrix tend to take on extreme values, not because they are correct but because they are subject to an extreme amount of error. Invariably the mean-variance optimization software will latch onto the extremes, and place its biggest bets on the coef-

Olivier Ledoit is a managing director in the Equities Division of Credit Suisse First Boston in London, UK.
olivier@ledoit.net

Michael Wolf is a an associate professor of economics and business at the Universitat Pompeu Fabra in Barcelona, Spain.
michael.wolf@upf.edu

ficients that are the most unreliable.

Michaud [1989] calls this phenomenon "error-maximization." It implies that managers' realized track records will underrepresent their true stock-picking abilities, which is clearly the last thing they would want.

Adding to this problem, some companies have proposed proprietary methods to generate covariance matrices that are advertised as better suited to mean-variance optimization than the sample covariance matrix. The drawbacks are that any manager using them establishes a costly and indefinite dependence on an external entity that does not share in any downside risk, and that the proprietary methods are not open to independent inspection and verification, so one can never be sure what is really going on behind the curtain.

We propose a new formula for estimating the covariance matrix of stock returns that is useful to replace the sample covariance matrix in any mean-variance optimization application, and it is absolutely free of charge and available to all. The method recognizes that the coefficients in the sample covariance matrix that are extremely high tend to be estimated with a lot of positive error, and therefore need to be pulled downward to compensate for that. Similarly, the method compensates for the negative error that tends to be embedded in extremely low estimated coefficients by pulling them upward. We call this the *shrinkage* of the extremes toward the center. If properly implemented, this shrinkage would clearly fix the problem of the sample covariance matrix.

The key questions are toward what target to shrink, and how intensely? Our contributions are: 1) to provide a rigorous statistical answer to both these questions; 2) to describe this so the reader can decide for himself or herself whether it makes sense; 3) to supply computer code that implements the resulting mathematical formula; and finally 4) to show that this approach produces significant improvements with actual stock return data.

Shrinkage is hardly a revolutionary concept in statistics, although it certainly was when it was first introduced by Stein [1955]. An excellent non-technical primer on shrinkage using real-life examples of baseball batting averages appears in Efron and Morris [1977]. That this idea has not yet percolated to a field where it would be most useful, portfolio management, is testimony to the Chinese walls between theoretical and applied disciplines whose adherents would benefit from talking to each other more. We endeavor to knock down these walls.

Early attempts to use shrinkage in portfolio selection were made by Frost and Savarino [1986] and Jorion [1986], but their particular shrinkage techniques were unable to cope with more stocks on the menu than in the historical return observations, as is very often the case. More recently, Jagannathan and Ma [2003] show that mean-variance optimizers already implicitly apply some form of shrinkage to the sample covariance matrix when short sales are ruled out, and that this is generally beneficial in terms of improving the stability of weights. All the more reason then to do this explicitly, so that the optimal shrinkage intensity can be applied.

We lay out much of the foundation for our work elsewhere (see Ledoit and Wolf [2003, 2004]). These are largely theoretical and general interest articles; here we focus specifically on how to employ the technology to add value to active portfolio management.

We first provide a formal description of the portfolio optimization problem in order to provide a base for our exposition. We then look at the out-of-sample performance of the estimator, using historical stock return data.

DESCRIPTION OF THE PROBLEM

We study the typical case for equity portfolios. The benchmark is a weighted index of a large number N of individual stocks, such as a value-weighted index. The universe of stocks from which the portfolio manager selects includes all these stocks. Excess returns are defined relative to the chosen benchmark.[1]

The notation is as follows:

w_B = vector of benchmark weights for the universe of N stocks;
x = vector of active weights;
$w_P = w_B + x$ = vector of portfolio weights;
y = vector of stock returns;
$\mu = \mathrm{E}(y)$ = vector of expected stock returns;
$\alpha = \mu - w'_B\mu$ = vector of expected stock excess returns; and
Σ = covariance matrix of stock returns.

We can write expected returns and variances in vector/matrix notation as:

$\mu_B = w'_B\mu$ = expected return on benchmark;
$\sigma_B^2 = w'_B\Sigma w_B$ = variance of benchmark return;
$\mu_P = w'_P\mu$ = expected return on portfolio;
$\sigma_P^2 = w'_P\Sigma w_P$ = variance of portfolio return;
$\mu_E = x'\mu$ = expected excess return on portfolio; and
$\sigma_E^2 = x'\Sigma x$ = tracking error variance.

The portfolio selection problem is subject to the constraint that the portfolio be fully invested; that is, the portfolio weights w_P must add up to unity. With **1** denoting a conforming vector of ones, this can be written as $w'_P \mathbf{1} = 1$. Because the benchmark weights also add to unity, the vector of portfolio deviations must add to zero, or $x' \mathbf{1} = 0$. Therefore, a manager's portfolio can be viewed as a position in the benchmark plus an *active portfolio*. The active portfolio is a long-short portfolio that expresses the views of the manager.

Two immediate implications are:

$$\mu_P = \mu_B + \mu_E$$

$$\sigma_P^2 = \sigma_B^2 + 2w'_B \Sigma x + \sigma_E^2$$

While positions of the active portfolio are both positive and negative, the manager does not have complete freedom. None of the portfolio weights w_P can be negative, or $w_P \geq 0$, due to the long-only constraint. The resulting constraint $x \geq -w_B$ expresses the limited freedom of the manager. Grinold and Kahn [2000, Chapter 15] illustrate how this limitation can negatively affect the performance of the managed portfolio, especially when the benchmark is a value-weighted index and when N is large. Another constraint is that the total position in any given stock cannot exceed a certain value, such as 10%. If this upper bound is denoted by c, the resulting constraint on the weights of the active portfolio is $x \leq c\mathbf{1} - w_B$.

We formalize the optimization problem of the manager as follows:[2]

$$\text{Minimize: } x'\Sigma x \tag{1}$$

$$\text{such that } x'\alpha \geq g$$
$$x'\mathbf{1} = 0$$
$$x' \geq -w_B$$
$$x \leq c\,\mathbf{1} - w_B$$

Here g is the manager's target gain (i.e., expected excess return) relative to the benchmark. A typical number is 300 basis points (annualized). The manager chooses g and the upper limit c and also knows the current vector of benchmark weights w_B. The manager is now left to provide estimates for α, the vector of expected stock excess returns, and for Σ, the covariance matrix of stock returns. In a final step, all the inputs are fed into a quadratic optimization software that will compute x, the optimal weights of the active portfolio.

We want to provide the manager with a good estimator of Σ. We do not address how to estimate α.

SHRINKAGE ESTIMATOR OF THE COVARIANCE MATRIX

The shrinkage estimator for Σ starts with the sample covariance matrix S. Its advantages are ease of computation and unbiasedness (i.e., its expected value is equal to the true covariance matrix). Its main disadvantage is that it is subject to a lot of estimation error when there are the same number of or even fewer data points than there are individual stocks; this is the typical case in financial applications.

Alternatively, one might consider an estimator with a lot of structure, like the single-factor model of Sharpe [1963]. Such estimators have relatively little estimation error but, on the other hand, they tend to be misspecified and can be severely biased. In one way or another, all successful risk models find a compromise between the sample covariance matrix and a highly structured estimator.

The industry standard is multifactor models. The idea is to incorporate multiple factors instead of just the single factor of Sharpe [1963], so the models become more flexible and their bias is reduced—but estimation error increases. Finding the optimal trade-off by deciding on the nature and the number of the factors included in the model is as much an art as it is a science.

One approach is to use a combination of *industry factors* and *risk indexes*, with factors on the order of 50. An example is Barra's U.S. equity model. Another approach is to use *statistical factors*, such as principal components, with factors on the order of 5. One commercial vendor offering risk models based on statistical factors is APT.

Our philosophy is different. Consider the sample covariance matrix S and a highly structured estimator, denoted by F. We find a compromise between the two by computing a convex linear combination $\delta F + (1 - \delta)S$, where δ is a number between 0 and 1. This technique is called *shrinkage*, as the sample covariance matrix is shrunk toward the structured estimator. The number δ is referred to as the *shrinkage constant*. Intuitively, it measures the weight that is given to the structured estimator.[3]

Shrinkage estimators have a long history in statistics. The beauty of the principle is that by properly combining two extreme estimators one can obtain a compromise estimator that performs better than either extreme. To draw a rough analogy: Most people would prefer the compromise of one bottle of Bordeaux and one steak to either

extreme of two bottles of Bordeaux (and no steak) or two steaks (and no Bordeaux).

Any shrinkage estimator has three ingredients: an estimator with no structure, an estimator with a lot of structure, and a shrinkage constant. The estimator without structure is generally quite obvious, given the context. For us it is the sample covariance matrix. Less obvious is the choice of the structured estimator, or shrinkage target, and the shrinkage constant.

Shrinkage Target

The shrinkage target should fulfill two requirements at the same time; it should involve only a small number of free parameters (that is, a lot of structure), but it should also reflect important characteristics of the unknown quantity to be estimated. In Ledoit and Wolf [2003] we suggest the single-factor matrix of Sharpe [1963] as the shrinkage target. Here we make a different suggestion: the *constant-correlation model*.

In our experience, the constant-correlation model gives comparable performance but is easier to implement. The model says that all the (pairwise) correlations are identical.[4]

Estimation of the model is straightforward. The average of all the sample correlations is the estimator of the common constant correlation. This number together with the vector of sample variances implies our shrinkage target, denoted by F henceforth. A formal description of the shrinkage target is provided in Appendix A; see in particular Equation (A).

Shrinkage Constant

The obvious practical problem is which value to choose for the shrinkage constant. Any choice of δ between 0 and 1 would yield a compromise between S and F, but this results in infinitely many possibilities. Intuitively, there is an optimal shrinkage constant, the one that minimizes the expected distance between the shrinkage estimator and the true covariance matrix. Call this number δ^*.

Appendix B derives a formula for estimating δ^*. The estimated optimal shrinkage constant is denoted $\hat{\delta}^*$; see Equation (B-2) in Appendix B. Our operational shrinkage estimator of the covariance matrix Σ is now:

$$\hat{\Sigma}_{Shrink} = \hat{\delta}^* F + (1 - \hat{\delta}^*) S \qquad (2)$$

EMPIRICAL STUDY

We describe the out-of-sample performance of our shrinkage estimator using historical stock market data. Datastream provides monthly U.S. stock data. We use these data to construct several value-weighted indexes to serve as benchmarks. Starting in February 1983, the methodology is as follows.

At the beginning of each month, we select the N largest stocks (with a ten-year history) as measured by their market value. The market values of the stocks define their index weights. At the end of the month, we observe the (real) returns of the individual stocks, and, given their weights, compute the return on the index. This prescription is repeated every month through December 2002 (yielding a total of 239 monthly returns). Thus, the constituent list and the index weights are constantly updated.

For benchmark size N, we use $N = 30$, 50, 100, 225, and 500. This range covers such important benchmarks as the DJIA, XETRA DAX, DJ STOXX 50, FTSE 100, Nasdaq-100, Nikkei 225, and S&P 500. Summary statistics of the various benchmark returns are provided in Exhibit 1. All numbers are annualized.

To mimic performance of a skilled active manager, we first construct *raw* forecasts of the expected excess returns by adding random noise to the realized excess returns. In a second step, these *raw* forecasts are transformed into *refined* forecasts $\hat{\alpha}$ that are fed to the quadratic optimizer. This is done so that the unconstrained annualized ex ante information ratio (IR) is approximately equal to 1.5, independently of the value of the benchmark size N. The unconstrained IR could be attained by a manager who did not face any lower or upper bound constraints on the weight vector x and who knew the exact nature of the covariance matrix Σ of stock returns.

The details of the forecast construction are described in Appendix C.

Evaluation Algorithm

A risk model is evaluated by its out-of-sample performance.

- At the beginning of each month, feed to the quadratic optimizer: the benchmark weights w_B, the forecasted expected excess returns $\hat{\alpha}$, the estimated covariance matrix $\hat{\Sigma}$, the desired gain g, and an upper bound of $c = 0.1$ on the total weight of any stock.

EXHIBIT 1
Summary Statistics for Benchmark Returns

	$N = 30$	$N = 50$	$N = 100$	$N = 225$	$N = 500$
Mean	13.63	13.50	13.29	13.45	13.42
Standard Deviation	15.12	15.02	14.76	14.56	14.52

- To compute an estimate $\hat{\Sigma}$, we use the last $T = 60$ monthly returns of the current constituent list of stocks.
- The quadratic optimizer computes a weight vector x.
- At the end of the month, the realized excess return is given by $e = x'\gamma$, where γ is the vector of stock returns for the month.
- The out-of-sample period ranges from February 1983 through December 2002, so we have a total of 239 monthly realized excess returns.
- From the excess returns we compute the (annualized) ex post information ratio as $\sqrt{12}\bar{e}/s_e$, where $\bar{e}$ is the sample average of the excess returns, and s_e is the sample standard deviation of the excess returns.
- Since the results depend on the monthly forecasts $\hat{\alpha}$, which are random, we repeat this process 50 times and then report mean summary statistics.
- For the (annualized) gain, we use 300 basis points.

Besides the shrinkage estimator, we include the sample covariance matrix, the shrinkage estimator of Ledoit and Wolf [2003], and a multifactor risk model based on statistical factors in our study.

The sample covariance matrix is widely known and very easy to compute. According to Jagannathan and Ma [2003], a portfolio manager facing a long-only constraint might hope it yields reasonable performance.

Multifactor models are the industry standard for estimation of the covariance matrix. Statistical factors, as opposed to macroeconomic and fundamental factors, have the advantage that they can be computed from past stock returns alone. The statistical factors we use are the first five principal components, as in Connor and Korajczyk [1993] and Connor [1995].[6]

Mean summary statistics for the realized excess returns are presented in Exhibit 2. Sample denotes the sample covariance matrix; shrink-CC denotes our shrinkage estimator; shrink-SF denotes the shrinkage estimator of Ledoit and Wolf [2003]; and PC-5 denotes the multifactor estimator based on the first five principal components. The results are mean summaries over 50 repetitions. All numbers are annualized.

A comparison of the shrinkage estimator we propose (shrink-CC) to the sample covariance matrix (sample) indicates that:

- In all scenarios, the shrinkage estimator yields the highest (average) information ratio.
- In most scenarios, the shrinkage estimator yields the highest (average) mean excess return.
- In all scenarios, the shrinkage estimator yields the lowest (average) standard deviation of excess return.
- The (average) information ratio declines as the benchmark size N increases.[7]

Exhibit 3 shows boxplots of the realized information ratios for the sample covariance matrix and the shrinkage estimator proposed here over the 50 repetitions for the various scenarios. For any given index size N, the first boxplot corresponds to the sample covariance matrix, and the second one to the shrinkage estimator in Equation (2). The message is identical to that in Exhibit 2: 1) shrinkage improves on the sample covariance matrix; and 2) realized information ratios tend to drop as N increases. The plots also show considerable variation in the realized information ratios. In addition to good forecasting skill and a good risk model, the successful active manager can benefit from a bit of good luck.

Shrink-CC does somewhat better than shrink-SF for $N \leq 100$ but somewhat worse for $N \geq 225$. Shrink-CC does somewhat better than PC-5 for $N \leq 50$, and is comparable to PC-5 for $N \geq 100$.

Exhibit 4 presents mean summary statistics on the average monthly turnover. Turnover is defined as *total turnover* following Grinold and Kahn [2000, Chapter 16]. Note that this means updating the entire portfolio, not just the active portfolio. A part of the turnover, therefore, is

EXHIBIT 2

Mean Summary Statistics for Excess Returns with Gain = 300 bp

	IR	Mean	SD
		$N = 30$	
Sample	0.97	2.18	2.26
Shrink-CC	1.24	2.50	2.03
Shrink-SF	1.18	2.39	2.04
PC-5	1.17	2.45	2.10
		$N = 50$	
Sample	0.79	1.92	2.44
Shrink-CC	1.14	2.21	1.95
Shrink-SF	1.08	2.13	1.98
PC-5	1.11	2.18	1.96
		$N = 100$	
Sample	0.59	1.71	2.93
Shrink-CC	0.91	1.87	2.06
Shrink-SF	0.89	1.86	2.10
PC-5	0.91	1.87	2.07
		$N = 225$	
Sample	0.37	2.37	6.45
Shrink-CC	0.54	2.53	4.97
Shrink-SF	0.57	2.37	4.30
PC-5	0.55	2.42	4.46
		$N = 500$	
Sample	0.20	1.92	8.53
Shrink-CC	0.30	1.82	5.77
Shrink-SF	0.33	1.74	5.13
PC-5	0.31	1.59	5.05

due to the constituent list of the benchmark and their weights, both of which change over time.

In general, the turnover is too high to be attractive for an active manager—but we made no effort to limit turnover, and a constraint on this could easily be added to the quadratic optimization in problem (1). The important message to take away from Exhibit 4 is that the sample covariance matrix always results in the highest turnover. The other three methods are comparable to one another.

More on Constraints

Many active managers face constraints besides the long-only constraint and an upper bound on the weight of any given stock. Examples are market cap-neutral constraints, turnover constraints, and dividend yield neutrality with respect to the benchmark. Adding further constraints generally reduces the ex post information ratio—see Clarke, de Silva, and Thorley [2002]—but the manager will still benefit from a superior risk model.

It is by now well understood that tracking error-efficient portfolios are not mean-variance efficient. The tracking error efficient frontier is shifted below and to the right of the mean-variance efficient frontier; see Roll [1992], Wilcox [1994], and Scherer [2002, Chapter 6]. Jorion [2003] shows that adding a constraint on the total portfolio variance, $\sigma_P^2 = w_P' \Sigma w_P$, to the quadratic optimization in problem (1) improved the mean-variance efficiency of the managed portfolio.[8]

Obviously, the additional constraint requires an estimate of Σ in practice, so a superior risk model will again be beneficial to the manager.

CONCLUSION

We propose a risk model that dominates the traditional one, the sample covariance matrix, for mean-variance optimization in the context of active portfolio management. Given the well-documented flaws of the sample covariance matrix, nobody should be using it now that an enhanced alternative is available. Using the simple modification we propose substantially increases the realized information ratio of the portfolio manager. If an annual expected excess return of 300 basis points over the benchmark is specified, a typical increase is on the order of 50%.

Computer code in the Matlab programming language implementing this improved estimator is freely downloadable.[9] Portfolio management firms that are sophisticated enough to use mean-variance optimization software would have the expertise required to implement our simple formulas in any computer language.

All types of portfolio optimization procedures, even advanced ones such as the resampled efficient frontier as in Michaud [1998] would benefit from shrinking the sample covariance matrix. The intuitive justification for this statistical transformation is prudence: not betting the ranch on noisy coefficients that are too extreme. We hope the profession can find value in our proposal.

APPENDIX A

Formula for Shrinkage Target

Let y_{it}, $1 \le i \le N$, $1 \le t \le T$, denote the return on stock i during period t. Our analysis assumes stock returns are independent and identically distributed (iid) over time and have finite fourth moments. The sample average of the returns of stock i is given by

EXHIBIT 3

Realized Information Ratios with Gain = 300 bp

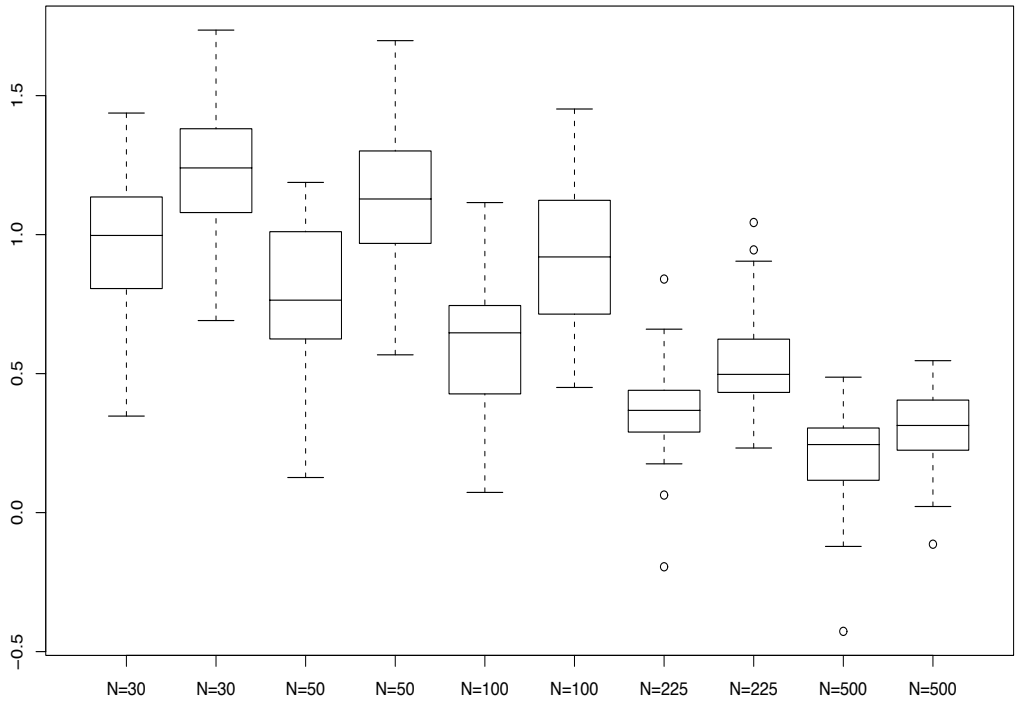

EXHIBIT 4

Mean Summary Statistics for Average Monthly Turnover

	$N = 30$	$N = 50$	$N = 100$	$N = 225$	$N = 500$
Sample	0.39	0.50	0.66	0.80	0.85
Shrink-CC	0.33	0.39	0.50	0.65	0.75
Shrink-SF	0.34	0.41	0.52	0.66	0.76
PC-5	0.33	0.39	0.50	0.64	0.73

$\bar{y}_{i\cdot} = T^{-1}\sum_{t=1}^{T} y_i$. Let $\mathbf{\Sigma}$ denote the population (or true) covariance matrix, and let S denote the sample covariance matrix. Typical entries of the matrices $\mathbf{\Sigma}$ and S are denoted by σ_{ij} and s_{ij}, respectively.

The population and sample correlations between the returns on stocks i and j are given by

$$\varrho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}$$

and

$$r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}s_{jj}}}$$

The average population and sample correlations are given by

$$\bar{\varrho} = \frac{2}{(N-1)N}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\varrho_{ij}$$

and

$$\bar{r} = \frac{2}{(N-1)N}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} r_{ij}$$

Define the population constant-correlation matrix $\mathbf{\Phi}$ by means of the population variances and the average population correlation:

$$\phi_{ii} = \sigma_{ii}$$

and

$$\phi_{ij} = \bar{\varrho}\sqrt{\sigma_{ii}\sigma_{jj}}$$

Correspondingly, define the sample constant-correlation matrix F by means of the sample variances and the average sample correlation:

$$f_{ii} = s_{ii}$$

and

$$f_{ij} = \bar{r}\sqrt{s_{ii}s_{jj}} \qquad \text{(A)}$$

APPENDIX B

Formula for Shrinkage Intensity

We have to choose the objective according to which the shrinkage intensity δ is optimal. All established shrinkage estimators from finite-sample statistical decision theory (also from Frost and Savarino [1986]) break down when $N \geq T$ because their loss functions involve the inverse of the covariance matrix. Instead, we propose a loss function that does not depend on this inverse and is very intuitive; it is a quadratic measure of distance between the true and the estimated covariance matrices based on the Frobenius norm.

The Frobenius norm of the $N \times N$ symmetric matrix Z with entries $(z_{ij})_{i,j=1,\ldots,N}$ is defined by

$$\|Z\|^2 = \sum_{i=1}^{N}\sum_{j=1}^{N} z_{ij}^2$$

By considering the Frobenius norm of the difference between the shrinkage estimator and the true covariance matrix, we arrive at the quadratic loss function:

$$\mathrm{L}(\delta) = \|\delta F + (1-\delta) S - \Sigma\|^2$$

The goal is to find the shrinkage constant δ that minimizes the expected value of this loss, that is, the risk:

$$\mathrm{R}(\delta) = \mathrm{E}(\mathrm{L}(\delta)) = \mathrm{E}\left(\|\delta F + (1-\delta) S - \Sigma\|^2\right) \quad \text{(B-1)}$$

Under the assumption that N is fixed while T tends to infinity, in Ledoit and Wolf [2003] we prove that the optimal value δ^* behaves asymptotically like a constant over T (up to higher-order terms). This constant, called κ, can be written as

$$\kappa = \frac{\pi - \rho}{\gamma}$$

where π denotes the sum of asymptotic variances of the entries of the sample covariance matrix scaled by $\sqrt{T}$:

$$\pi = \sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{AsyVar}\left[\sqrt{T} s_{ij}\right]$$

Similarly, ρ denotes the sum of asymptotic covariances of the entries of the shrinkage target with the entries of the sample covariance matrix scaled by $\sqrt{T}$:

$$\rho = \sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{AsyCov}\left[\sqrt{T} f_{ij}, \sqrt{T} s_{ij}\right]$$

Finally, γ measures the misspecification of the (population) shrinkage target:

$$\gamma = \sum_{i=1}^{N}\sum_{j=1}^{N} (\phi_{ij} - \sigma_{ij})^2$$

If κ were known, we could use κ/T as the shrinkage intensity in practice. Unfortunately, κ is unknown, so we find a consistent estimator for κ. This is done by finding consistent estimators for the three ingredients π, ρ, and γ.

First, a consistent estimator for π is

$$\hat{\pi} = \sum_{i=1}^{N}\sum_{j=1}^{N} \hat{\pi}_{ij}$$

with

$$\hat{\pi}_{ij} = \frac{1}{T}\sum_{t=1}^{T} \left\{(y_{it} - \bar{y}_{i\cdot})(y_{jt} - \bar{y}_{j\cdot}) - s_{ij}\right\}^2$$

This result is proven by Ledoit and Wolf [2003].

Second, a consistent estimator for ρ is a bit tedious to write but quite straightforward to implement. By definition:

$$\begin{aligned}\rho &= \sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{AsyCov}\left[\sqrt{T} f_{ij}, \sqrt{T} s_{ij}\right] \\ &= \sum_{i=1}^{N} \mathrm{AsyVar}\left[\sqrt{T} s_{ii}\right] + \\ &\quad \sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} \mathrm{AsyCov}\left[\sqrt{T}\,\bar{r}\sqrt{s_{ii}s_{jj}}, \sqrt{T} s_{ij}\right]\end{aligned}$$

On the diagonal, we know from Ledoit and Wolf [2003] again that:

$$\mathrm{AsyVar}\left[\sqrt{T} s_{ii}\right] = \hat{\pi}_{ii} = \frac{1}{T}\sum_{t=1}^{T} \left\{(y_{it} - \bar{y}_{i\cdot})^2 - s_{ii}\right\}^2$$

On the off-diagonal, we have:

$$\mathrm{AsyCov}\left[\sqrt{T}\,\bar{r}\sqrt{s_{ii}s_{jj}}, \sqrt{T} s_{ij}\right]$$

Since the estimation error in $\bar{r}$ is asymptotically negligible and by use of the delta method, any term

$$\mathrm{AsyCov}\left[\sqrt{T}\,\bar{r}\sqrt{s_{ii}s_{jj}}, \sqrt{T} s_{ij}\right]$$

can be consistently estimated by

$$\frac{\bar{r}}{2}\left(\sqrt{\frac{s_{jj}}{s_{ii}}}\,\mathrm{AsyCov}\left[\sqrt{T} s_{ii}, \sqrt{T} s_{ij}\right] + \sqrt{\frac{s_{ii}}{s_{jj}}}\,\mathrm{AsyCov}\left[\sqrt{T} s_{jj}, \sqrt{T} s_{ij}\right]\right)$$

Standard theory implies that a consistent estimator for AsyCov $[\sqrt{T} s_{ii}, \sqrt{T} s_{ij}]$ is given by

$$\hat{\vartheta}_{ii,ij} = \frac{1}{T}\sum_{t=1}^{T}\left\{(y_{it}-\bar{y}_{i\cdot})^2 - s_{ii}\right\}\times$$

$$\left\{(y_{it}-\bar{y}_{i\cdot})(y_{jt}-\bar{y}_{j\cdot}) - s_{ij}\right\}$$

and that, analogously, a consistent estimator for AsyCov $[\sqrt{T}s_{ii}, \sqrt{T}s_{ij}]$ is given by:

$$\hat{\vartheta}_{jj,ij} = \frac{1}{T}\sum_{t=1}^{T}\left\{(y_{jt}-\bar{y}_{j\cdot})^2 - s_{jj}\right\}\times \left\{(y_{it}-\bar{y}_{i\cdot})(y_{jt}-\bar{y}_{j\cdot}) - s_{ij}\right\}$$

Collecting terms now yields a consistent estimator for ρ:

$$\hat{\rho} = \sum_{i=1}^{N}\hat{\pi}_{ii} + \sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\frac{\bar{r}}{2}\left(\sqrt{\frac{s_{jj}}{s_{ii}}}\,\hat{\vartheta}_{ii,ij} + \sqrt{\frac{s_{ii}}{s_{jj}}}\,\hat{\vartheta}_{jj,ij}\right)$$

Third, a consistent estimator for γ is

$$\hat{\gamma} = \sum_{i=1}^{N}\sum_{j=1}^{N}(f_{ij}-s_{ij})^2$$

This result follows because f_{ij} and s_{ij} are consistent estimators of ϕ_{ij} and σ_{ij}, respectively.

Putting the pieces together yields a consistent estimator for κ:

$$\hat{\kappa} = \frac{\hat{\pi}-\hat{\rho}}{\hat{\gamma}}$$

Finally, the estimated shrinkage intensity we propose for use in practice is:

$$\hat{\delta}^* = \max\left\{0, \min\left\{\frac{\hat{\kappa}}{T}, 1\right\}\right\} \qquad \text{(B-2)}$$

Although it is very unlikely, in principle it can happen in a finite sample that $\hat{\kappa}/T < 0$, or that $\hat{\kappa}/T > 1$, in which case we simply truncate it at 0 or at 1, respectively.

APPENDIX C

Forecasting Expected Excess Returns

We want to mimic a skilled active manager. To do this, we rely on hindsight and the forecast principles laid out in Grinold and Kahn [2000, Chapters 6 and 10].

Let e_{it} denote the excess return of stock i during period t, that is, the return on the stock minus the benchmark return. In a first step, we generate *raw forecasts* by adding noise to the realized excess returns:

$$raw_{it} = e_{it} + u_{it}$$

The noise terms u_{it} are normally distributed with mean zero and are independent of each other both cross-sectionally and over time. For a given stock, the ex ante correlation between e_{it} and u_{it} over time is known as the information coefficient (IC). In principal, the IC could depend on i but, as is common, we choose a coefficient constant across stocks.

What is an appropriate value for the IC? In the absence of constraints on the active manager (apart from being fully invested), the well-known fundamental law of active management states that the ex ante information ratio (IR) of the manager is determined by the IC and the breadth of the strategy:

$$\text{IR} \approx \text{IC}\sqrt{\text{Breadth}}$$

The breadth term measures the number of independent active bets the manager makes in one year. Since in our study the portfolio will be updated every month and, by construction, the forecasts are independent of each other, we have

$$\text{Breadth} = 12N$$

Therefore, the IC is determined by the size of the benchmark, N, and the desired ex ante information ratio, IR, which we fix at 1.5. Putting the various pieces together yields:

$$\text{IC} = 1.5/\sqrt{12N}$$

To give three examples, $N = 30$ yields IC = 0.0791, $N = 100$ yields IC = 0.0433, and $N = 500$ yields IC = 0.0194.

In a second step, the raw forecasts for each stock are converted to *scores* by subtracting their sample mean, $\overline{raw}_{i\cdot}$, and dividing by their sample standard deviation, $s_{raw,i}$:

$$score_{it} = \frac{raw_{it} - \overline{raw}_{i\cdot}}{s_{raw,i}}$$

In a third and final step, the scores are transformed into *refined forecasts* using the relationship:

$$\text{Alpha} = \text{Volatility} \times \text{IC} \times \text{Score}$$

Here, volatility refers to the excess return of a given stock. We estimate this by the sample standard deviation of the realized excess returns e_{it} over time. Denoting this standard deviation by $s_{e,i}$, the formula for the final step is:

$$\hat{\alpha}_{it} = s_{e,i} \times \text{IC} \times score_{it}$$

Note that the ex post information ratio of an active manager will in general not be equal to the ex ante value of 1.5. This is because 1) the manager is bound by constraints (such as a long-only constraint and upper limits on the portfolio weight of each stock); 2) the manager has to estimate $\boldsymbol{\Sigma}$ in practice; 3) with the randomness of the u_{it}, the ex post correlation between e_{it} and u_{it} over time will not be equal to IC.

The first two facts have a negative effect on the ex post information ratio. The third fact can go either way. It is therefore possible in practice, although not very likely, that the ex post information ratio is higher than the ex ante value of 1.5. To smooth the inherent randomness in the realized information ratios, we repeat the forecasting process 50 times and then report mean summaries.

ENDNOTES

[1]The problem can be generalized to a universe that includes stocks not in the benchmark. If we want to minimize transaction costs, however, the more general setting is of limited practical interest.

[2]Jorion [2003] considers the problem of maximizing $x'\alpha$ subject to an upper bound on the tracking error variance $x'\Sigma x$. Grinold and Kahn [2000] consider the problem of maximizing $x'\alpha - \lambda x'\Sigma x$, where λ is a risk aversion constant. These are equivalent problem formulations, leading to the same frontier in risk-return space.

[3]In Ledoit and Wolf [2003] we denote the shrinkage constant by α. We use the symbol δ here to avoid confusion with expected excess returns.

[4]The constant-correlation model would not be appropriate if the assets are from different *asset classes,* such as stocks and bonds. In such cases, more general models for the shrinkage target are available.

[5]An additional advantage of our shrinkage estimator is that it is always positive-definite. When the size of the benchmark N exceeds the number of past returns used in the estimation process, the sample covariance matrix S is singular. This would imply that there are stock portfolios with zero risk. Our shrinkage estimator, however, is a convex combination of an estimator that is positive-definite (the shrinkage target F) and an estimator that is positive semidefinite (the sample covariance matrix S). It is therefore positive-definite.

[6]Multifactor models based on statistical factors require more programming effort than our shrinkage estimator. First, one needs software to extract the factors, such as principal components, from the data. Second, various time series regressions have to be run to estimate the factor loadings and residual variances. Third, the various pieces have to be put together to arrive at the estimator of the covariance matrix.

[7]Grinold and Kahn [2000, Chapter 15] explain why this happens. If N is very large, a manager is probably ill-advised to actively invest in all the stocks making up the index. The realized information ratio can be improved by, for example, focusing on the 50 or 100 largest stocks in the index and setting the weights of the remaining stocks equal to zero.

[8]A related idea appeared first in Wilcox [1994].

[9]The address is http://www.ledoit.net.

REFERENCES

Clarke, Roger, Harindra de Silva, and Steven Thorley. "Portfolio Constraints and the Fundamental Law of Active Management." *Financial Analysts Journal,* 58 (2002), pp. 48-66.

Connor, Gregory. "The Three Types of Factor Models: A Comparison of Their Explanatory Power." *Financial Analysts Journal,* 51 (1995), pp. 42-46.

Connor, Gregory, and Robert A. Korajczyk. "A Test for the Number of Factors in an Approximate Factor Model." *Journal of Finance,* 48 (1993), pp. 1263-1291.

Efron, Bradley, and Carl Morris. "Stein's Paradox in Statistics." *Scientific American,* 236 (1977), pp. 119-127.

Frost, Peter A., and James E. Savarino. "An Empirical Bayes Approach to Portfolio Selection." *Journal of Financial and Quantitative Analysis,* 21 (1986), pp. 293-305.

Grinold, Richard C., and Ronald N. Kahn. *Active Portfolio Management,* 2nd ed. New York: McGraw-Hill, 2000.

Jagannathan, Ravi, and Tongshu Ma. "Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps." *Journal of Finance,* 54 (2003), pp. 1651-1684.

Jobson, J. Dave, and Bob M. Korkie. "Estimation for Markowitz Efficient Portfolios." *Journal of the American Statistical Association*, 75 (1980), pp. 544-554.

Jorion, Philippe. "Bayes-Stein Estimation for Portfolio Analysis." *Journal of Financial and Quantitative Analysis*, 21 (1986), pp. 279-292.

——. "Portfolio Optimization with Constraints on Tracking Error." *Financial Analysts Journal*, 59 (2003), pp. 70-82.

Ledoit, Olivier, and Michael Wolf. "Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection." *Journal of Empirical Finance,* 10 (2003), pp. 603-621.

——. "A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices." *Journal of Multivariate Analysis*, 88 (2004), 365-411.

Markowitz, Harry. "Portfolio Selection." *Journal of Finance*, 7 (1952), pp. 77-91.

Michaud, Richard. *Efficient Asset Management: A Practical Guide to Stock Portfolio Optimization*. New York: Oxford University Press, 1998.

——. "The Markowitz Optimization Enigma: Is Optimized Optimal?" *Financial Analysts Journal,* 45 (1989), pp. 31-42.

Roll, Richard. "A Mean/Variance Analysis of Tracking Error." *The Journal of Portfolio Management,* 18 (1992), pp. 13-22.

Scherer, Bernd. *Portfolio Construction and Risk Budgeting*. London: Risk Books, 2002.

Sharpe, William F. "A Simplified Model for Portfolio Analysis." *Management Science,* 9 (1963), pp. 277-293.

Stein, Charles. "Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution." In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley: University of California Press, 1955, pp. 197-206.

Wilcox, Jarrod. "EAFE is for Wimps." *The Journal of Portfolio Management*, 20 (1994), pp. 68-75.