# Honey, I Shrunk the Sample Covariance Matrix

Olivier Ledoit

Equities Division

Credit Suisse First Boston

One Cabot Square

London E14 4QJ, UK

olivier@ledoit.net

Michael Wolf*

Department of Economics and Business

Universitat Pompeu Fabra

Ramon Trias Fargas, 25–27

08005 Barcelona, Spain

michael.wolf@upf.edu

November 2003

## Abstract

The central message of this paper is that nobody should be using the sample covariance matrix for the purpose of portfolio optimization. It contains estimation error of the kind most likely to perturb a mean-variance optimizer. In its place, we suggest using the matrix obtained from the sample covariance matrix through a transformation called *shrinkage*. This tends to pull the most extreme coefficients towards more central values, thereby systematically reducing estimation error where it matters most. Statistically, the challenge is to know the optimal shrinkage intensity, and we give the formula for that. Without changing any other step in the portfolio optimization process, we show on actual stock market data that shrinkage reduces tracking error relative to a benchmark index, and substantially increases the realized information ratio of the active portfolio manager.

*Phone: +34-93-542 2552, Fax: +34-93-542 1746

# 1  Introduction

Since the seminal work of Markowitz (1952), mean-variance optimization has been the most rigorous way to pick stocks in which to invest. 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 forbidden from selling any stock short, and in the modern world they are typically measured against the benchmark of an equity market index with fixed (or infrequently rebalanced) weights. Fast and accurate quadratic optimization softwares exist that can solve this problem — provided they are fed the right inputs, that is.

Estimating the covariance matrix of stock returns has always been one of the stickiest points. The standard statistical method is to gather a history of past stock returns and compute their *sample covariance matrix*. Unfortunately this creates problems that are well documented (Jobson and Korkie, 1980). To put it as simply as possible, when the number of stocks under consideration is large, especially relative to the number of historical return observations available (which is the usual case), the sample covariance matrix is estimated with a lot of error. It implies that the most extreme coefficients in the matrix thus estimated tend to take on extreme values not because this is "the truth", but because they contain an extreme amount of error. Invariably the mean-variance optimization software will latch onto them and place its biggest bets on those coefficients which are the most extremely 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 want.

On the back of this, some companies such as APT and BARRA have proposed proprietary methods to generate covariance matrices that are advertized 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 who does not share in any downside risk, and that their proprietary methods are not open for independent inspection and verification, so one can never be sure what is really going on behind the curtain.

This is why we propose a new formula for estimating the covariance matrix of stock returns that can beneficially replace the sample covariance matrix in any mean-variance optimization application, and is absolutely free of charge and open to everybody. The crux of the method is that those estimated coefficients in the sample covariance matrix that are extremely high tend to contain a lot of positive error and therefore need to be pulled downwards to compensate for that. Similarly, we compensate for the negative error that tends to be embedded inside extremely low estimated coefficients by pulling

them upwards. We call this the *shrinkage* of the extremes towards the center. If properly implemented, this shrinkage would clearly fix the problem of the sample covariance matrix described above. The key questions are towards 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 below so the reader can decide for him or herself whether it makes sense; (3) to supply computer code that implements the resulting mathematical formula; and finally (4) to show that it yields significant improvements on actual stock return data.

Shrinkage is hardly a new and revolutionary concept in Statistics, although it certainly was when first introduced by Professor Charles Stein of Stanford University in 1955. An excellent non-technical primer on shrinkage using real-life examples of baseball batting averages was written by Efron and Morris (1977). That this idea has not yet percolated to a field where it would be most useful, portfolio management, is a testimony that Chinese walls still exist between theoretical and applied disciplines who 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 broke down when the number of stocks on the menu exceeds the number of historical return observations, which is very often the case in practice. Recently Jagannathan and Ma (2003) proved that mean-variance optimizers are already implicitly applying 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 weights stability. All the more reason then to do it explicitly so that the optimal shrinkage intensity can be applied, as in our paper. Much of the foundations for the present work has been laid out by the authors in other papers (Ledoit and Wolf, 2003, 2004). Those were largely theoretical and general-interest articles, whereas the present one focuses specifically on how to employ the technology to add value to active portfolio management.

First we give a formal description of the portfolio optimization problem in order to provide a solid base for our later exposition. Having described our estimator, we then look at its out-of-sample performance, using historical stock return data.

## 2    Formal Description of the Problem

We study the most relevant 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.[1]  Excess

returns are defined relative to the chosen benchmark. Define the following notations:

$$
\begin{aligned}
w_B &= \text{ vector of benchmark weights for the universe of } N \text{ stocks} \\
x &= \text{ vector of active weights} \\
w_P = w_B + x &= \text{ vector of portfolio weights} \\
y &= \text{ vector of stock returns} \\
\mu = \text{E}(y) &= \text{ vector of expected stock returns} \\
\alpha = \mu - w_B'\mu &= \text{ vector of expected stock excess returns} \\
\Sigma &= \text{ covariance matrix of stock returns}
\end{aligned}
$$

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

$$
\begin{aligned}
\mu_B = w_B'\mu &= \text{ expected return on benchmark} \\
\sigma_B^2 = w_B'\Sigma w_B &= \text{ variance of benchmark return} \\
\mu_P = w_P'\mu &= \text{ expected return on portfolio} \\
\sigma_P^2 = w_P'\Sigma w_P &= \text{ variance of portfolio return} \\
\mu_E = x'\mu &= \text{ expected excess return on portfolio} \\
\sigma_E^2 = x'\Sigma x &= \text{ tracking error variance}
\end{aligned}
$$

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

$$
\begin{aligned}
\mu_P &= \mu_B + \mu_E \\
\sigma_P^2 &= \sigma_B^2 + 2w_B'\Sigma x + \sigma_E^2
\end{aligned}
$$

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. In addition, the manager often is faced with the constraint that the total position in any given stock

cannot exceed a certain value, like 10%. If this upper bound is denoted by $c$, the resulting constraint on the weights of the active portfolio is $x \leq c\mathbf{1} - w_B$.

Having defined the various ingredients, we can now formalize the optimization problem of the manager as follows:[2]

$$
\begin{aligned}
\text{Minimize:} \quad & x'\Sigma x & (1)\\
\text{such that:} \quad & x'\alpha \geq g \\
& x'\mathbf{1} = 0 \\
& x \geq -w_B \\
& x \leq c\mathbf{1} - w_B
\end{aligned}
$$

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 $w_B$. She is now left to provide estimates for $\alpha$, the vector of expected stock excess returns, and for $\Sigma$, 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. It is our mission to provide the manager with a good estimator of $\Sigma$. We do not help with the problem of how to estimate $\alpha$.

# 3   Shrinkage Estimator of the Covariance Matrix

## 3.1   Shrinkage Principle

This section briefly describes the shrinkage estimator for $\Sigma$ which we propose; for a more detailed exposition and relevant historical background the reader is referred to Ledoit and Wolf (2003). We start with the sample covariance matrix $S$. Its advantages are ease of computation and the property of being unbiased (i.e., its expected value is equal to the true covariance matrix). Its main disadvantage is the fact that it contains a lot of estimation error when the number of data points is of comparable or even smaller order than the number of individual stocks; this is the common situation in financial applications. Alternatively, one might consider an estimator with a lot of structure, as the single-factor model of Sharpe (1963). Such estimators contain relatively little estimation error but, on the other hand, 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 are multi-factor models. The idea is to incorporate multiple

factors instead of just the single factor of Sharpe (1963). Thereby the models become more flexible and their bias is reduced. But the estimation error increases. Finding the optimal tradeoff 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 indices*, with the total number of factors being 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 the total number of factors being on the order of 5. A 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 $\delta$ is a number between 0 and 1. This technique is called *shrinkage*, since the sample covariance matrix is 'shrunk' towards the structured estimator. The number $\delta$ is referred to as the *shrinkage constant*.[3] Intuitively, it measures the weight that is given to the structured estimator. Shrinkage estimators have a long and successful 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 make a somewhat sloppy analogy: most people would be 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 are the choice of the structured estimator, or shrinkage target, and the shrinkage constant.

## 3.2   Shrinkage Target

The shrinkage target should fulfill two requirements at the same time: it involves only a small number of free parameters (that is, a lot of structure) but it also reflects important characteristics of the unknown quantity being estimated. Ledoit and Wolf (2003) suggest the single-factor matrix of Sharpe (1963) as the shrinkage target. In this paper we make a different suggestion: the *constant correlation model*. In our experience, it gives comparable performance but is easier to implement. The model says that all the (pairwise) correlations are identical.[4] The 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 $F$ in the remainder of the paper. A formal description of the shrinkage target

is provided in Appendix A; in particular, see equation (3).

## 3.3  Shrinkage Constant

The obvious practical problem is which value to choose for the shrinkage constant. Any choice of $\delta$ strictly 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. It is the one that minimizes the expected distance between the shrinkage estimator and the true covariance matrix. Call this number $\delta^*$. Appendix B derives a formula for estimating $\delta^*$. The estimated optimal shrinkage constant is denoted $\hat{\delta}^*$; see equation (5) in Appendix B. Our operational shrinkage estimator of the covariance matrix $\Sigma$ is now ready for use[5]:

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

# 4  Empirical Study

We now study 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 indices to serve as our benchmarks. Starting in February 1983, the methodology is as follows. At the beginning of each month, we select the $N$ largest stocks as measured by their market value. The market values of the stocks define their index weight. 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 until the end of December 2002. Thus, the constituents list and the index weights are constantly updated.

As far as the benchmark size $N$ is concerned, we employ $N = 30, 50, 100, 225, 500$. This range covers such important benchmarks as 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 Table 1.

To mimic 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 $\hat{\alpha}$ which are fed to the quadratic optimizer. This is done in a way such 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 $\Sigma$ of stock returns. The details of the forecast construction are described in Appendix C.

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

**Evaluation Algorithm:**

- At the beginning of each month feed the following ingredients 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.

- To compute an estimate $\hat{\Sigma}$, we use the last $T = 60$ monthly returns of the current constituents list of stocks.

- The quadratic optimizer computes a weight vector $x$.

- At the end of the month, the realized excess return is given by $e = x'y$, where $y$ is the vector of stock returns for the month.

- The out-of-sample period ranges from 02/1983 until 12/2002, so a total of 239 monthly realized excess returns are obtained.

- 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.

Apart from our shrinkage estimator, we include the sample covariance matrix, the shrinkage estimator of Ledoit and Wolf (2003), and a multi-factor risk model based on statistical factors in our study.

The sample covariance matrix is widely-known and very easy to compute. Based on the recent paper by Jagannathan and Ma (2003), a portfolio manager facing a long-only constraint might hope that it yields reasonable performance.

Ledoit and Wolf (2003) suggest the single-factor model of Sharpe (1963) as the shrinkage target, instead of the constant correlation covariance matrix as we do in this paper.

Multi-factor models are the industry standard for the 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.[6] The statistical factors we use are the first five principal components, as in Connor and Korajczyk (1993) and Connor (1995).

Mean-summary statistics for the realized excess returns are presented in Table 2. Comparing the shrinkage estimator proposed in this paper (Shrink-CC) to the sample covariance matrix (Sample), the results can be highlighted as follows.

- 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 decreases as the benchmark size $N$ increases.[7]

Figure 1 shows boxplots of the realized information ratios for the sample covariance matrix and the shrinkage estimator proposed in this paper over the 50 repetitions for the various scenarios. The message is identical to the one of Table 2: (i) shrinkage improves on the sample covariance matrix; (ii) realized information ratios tend to decrease as $N$ increases. The plots also show considerable variation in the realized information ratios. In addition to having good forecasting skill and using a good risk model, the successful active manager can benefit from a bit of good luck.

Next, we compare the shrinkage estimator proposed in this paper (Shrink-CC) to the other methods included in the study. The single-factor shrinkage estimator proposed by Ledoit and Wolf (2003) is denoted by Shrink-SF and the 5 principal components covariance matrix estimator is denoted by PC-5. Shrink-CC does somewhat better than Shrink-SF for $N \leq 100$ and is 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 $N \geq 100$.

Table 3 presents mean-summary statistics on the average monthly turnover. Turnover is defined as the *total turnover* of Grinold and Kahn (2000, Chapter 16) Note that this definition corresponds to updating the entire portfolio, not just the active portfolio. A part of the turnover, therefore, is due to the constituents 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 no effort was made to limit turnover, and a constraint to this effect could be easily added to the quadratic optimization problem (1). The important message to take away from Table 3 is that the sample covariance matrix

results in the highest turnover always. The other three methods are comparable to one another.

We finish this section with two remarks.

**Remark 1** Many active managers face further constraints, apart from 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 et al. (2002). Nevertheless, the manager will still benefit from using a superior risk model.

**Remark 2** 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 compared to the mean-variance efficient frontier. For example, 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 problem (1) improves the mean-variance efficiency of the managed portfolio.[8] Obviously, the additional constraint requires an estimate of $\Sigma$ in practice and therefore a superior risk model will again be beneficial to the manager.

# 5   Conclusion

This paper proposes a risk model that dominates the traditional one, the sample covariance matrix, for the purpose of mean-variance optimization in the context of active portfolio management. We claim that, given the well-documented flaws of the sample covariance matrix, nobody should be using it anymore now that an enhanced alternative is available. Using the simple modification we propose substantially increases the realized information ratio of the portfolio manager. For example, when 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 from `http://www.ledoit.net`. Portfolio management firms that are sophisticated enough to employ a 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[9], 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 that the profession can find value in our proposal.

# Notes

[1]The problem can be generalized to the setting where the universe contains further stocks not contained in the benchmark. However, to keep transaction costs down, 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 $\lambda$ is a risk-aversion constant. These are equivalent problem formulations, leading to the same frontier in risk-return space.

[3]Ledoit and Wolf (2003) denote the shrinkage constant by $\alpha$. We switch to the symbol $\delta$ in this paper to avoid confusion with expected excess returns.

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

[5]An additional advantage of our srhinkage 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 threre exist stock portfolios with zero risk. On the other hand, our shinkage estimator is a convex combination of an estimator that is positive definite (the shrinkage target $F$) and an estimator that is positive semi-definite (the sample covariance matrix $S$). It is therefore positive definite.

[6]Nevertheless, multi-factor 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, residual variances, etc.. 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. In case $N$ is very large, a manager is probably ill-advised to actively invest in all the stocks making up the index. The realized information ration can be improved by, for example, focusing on the 50 or 100 largest stocks in the index and setting the weights of the remaining ones equal to zero.

[8]A related idea already appears in Wilcox (1994).

[9]Future research will focus on the effectiveness of resampling methods in the spirit of Michaud (1998), either by themselves or combined with shrinkage, compared to the shrinkage method proposed in this paper.

# A    Formula for Shrinkage Target

Some notation is needed. Let $y_{it}$, $1 \leq i \leq N, 1 \leq t \leq T$, denote the return on stock $i$ during period $t$. Our analysis assumes that 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 $\bar{y}_{i\cdot} = T^{-1} \sum_{t=1}^{T} y_{it}$. Let $\Sigma$ denote the population (or true) covariance matrix and let $S$ denote the sample covariance matrix. Typical entries of the matrices $\Sigma$ and $S$ are denoted by $\sigma_{ij}$ and $s_{ij}$, respectively.

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

$$\varrho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}} \qquad \text{and} \qquad r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}s_{jj}}}$$

The average population and sample correlations are given by

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

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

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

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

$$f_{ii} = s_{ii} \qquad \text{and} \qquad f_{ij} = \bar{r}\sqrt{s_{ii}s_{jj}} \tag{3}$$

This matrix $F$ is the shrinkage target introduced in Subsection 3.2.

# B    Formula for Shrinkage Intensity

We have to choose the objective according to which the shrinkage intensity $\delta$ is optimal. All existing shrinkage estimators from finite-sample statistical decision theory and also the one of 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 $(z_{ij})_{i,j=1,\dots,N}$ 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 following quadratic loss function:

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

The goal is to find the shrinkage constant $\delta$ which minimizes the expected value of this loss, that is, the risk:

$$\mathrm{R}(\delta) = \mathrm{E}(\mathrm{L}(\delta)) = \mathrm{E}\left(\|\delta\, F + (1 - \delta)\, S - \Sigma\|^2\right) \tag{4}$$

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

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

Here, $\pi$ 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, $\rho$ 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, $\gamma$ measures the misspecification of the (population) shrinkage target: $\gamma = \sum_{i=1}^{N} \sum_{j=1}^{N} (\phi_{ij} - \sigma_{ij})^2$.

If $\kappa$ were known, we could use $\kappa/T$ as the shrinkage intensity in practice. Unfortunately, $\kappa$ is unknown, so we find a consistent estimator for $\kappa$. This is done by finding consistent estimators for the three ingredients $\pi$, $\rho$, and $\gamma$.

First, a consistent estimator for $\pi$ is

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

This result is proven by Ledoit and Wolf (2003).

Second, a consistent estimator for $\rho$ is a bit tedious to write down but quite straight-

forward to implement. By definition

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

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 $\mathrm{AsyCov}\left[\sqrt{T} s_{ii}, \sqrt{T} s_{ij}\right]$ is given by

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

and that, analogously, a consistent estimator for $\mathrm{AsyCov}\left[\sqrt{T} s_{jj}, \sqrt{T} s_{ij}\right]$ is given by

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

Collecting terms now yields a consistent estimator for $\rho$:

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

Third, a consistent estimator for $\gamma$ is

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

This result follows from $f_{ij}$ and $s_{ij}$ being consistent estimators of $\phi_{ij}$ and $\sigma_{ij}$, respectively.

Putting the pieces together yields a consistent estimator for $\kappa$:

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

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

$$\hat{\delta}^* = \max\left\{0, \min\left\{\frac{\hat{\kappa}}{T}, 1\right\}\right\} \tag{5}$$

The reason for this formula is the following. Although very unlikely, in principle it can happen in 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.

# 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). Our description is necessarily brief and the reader is referred to this book for further details.

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. Which is an appropriate value for 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:

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

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

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

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

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

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, $\overline{raw}_{i.}$, and dividing by their sample standard deviation, $s_{raw,i}$.

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

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

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

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

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

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 (i) the manager is bound by constraints (such as a long-only constraint and upper limits on the portfolio weight of each stock); (ii) the manager has to estimate $\Sigma$ in practice; (iii) due to the randomness of the $u_{it}$, the *ex post* correlation between 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, though not very likely, that the *ex post* information ratio is higher than the *ex ante value* of 1.5. To smooth out the

inherent randomness in the realized information ratios, we repeat the forecasting process 50 times in our empirical study of Section 4 and then report mean-summaries.

# References

Clarke, R., de Silva, H., and Thorley, S. (2002). Portfolio constraints and the fundamental law of active management. *Financial Analysts Journal*, 58(5):48–66.

Connor, G. (1995). The three types of factor models: A comparison of their explanatory power. *Financial Analysts Journal*, 51(3):42–46.

Connor, G. and Korajczyk, R. A. (1993). A test for the number of factors in an approximate factor model. *Journal of Finance*, 48:1263–1291.

Efron, B. and Morris, C. (1977). Stein's paradox in statistics. *Scientific American*, 236(5):119–127.

Frost, P. A. and Savarino, J. E. (1986). An empirical Bayes approach to portfolio selection. *Journal of Financial and Quantitative Analysis*, 21:293–305.

Grinold, R. C. and Kahn, R. N. (2000). *Active Portfolio Management.* McGraw-Hill, New York, second edition.

Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance*, 54(4):1651–1684.

Jobson, J. D. and Korkie, B. (1980). Estimation for Markowitz efficient portfolios. *Journal of the American Statistical Association*, 75:544–554.

Jorion, P. (1986). Bayes-Stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis*, 21:279–292.

Jorion, P. (2003). Portfolio optimization with constraints on tracking error. *Financial Analysts Journal*, 59(5):70–82.

Ledoit, O. and Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, 10(5):603–621.

Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, 7:77–91.

Michaud, R. (1989). The Markowitz optimization enigma: Is optimized optimal? *Financial Analysts Journal*, 45:31–42.

Michaud, R. (1998). *Efficient Assset Management: A Practial Guide to Stock Portfolio Optimization*. Oxford University Press.

Roll, R. (1992). A mean/variance analysis of tracking error. *Journal of Portfolio Management*, 18(4):13–22.

Scherer, B. (2002). *Portolio Construction and Risk Budgeting*. Risk Books, London.

Sharpe, W. F. (1963). A simplified model for portfolio analysis. *Management Science*, 9(1):277–293.

Stein, C. (1955). 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*, pages 197–206. University of California Press.

Wilcox, J. (1994). EAFE is for wimps. *Journal of Portfolio Management*, 20:68–75.

# D    Tables and Figures

Table 1: **Summary Statistics of Benchmark Returns.** This table presents summary statistics for monthly real returns of several value-weighted benchmark indices. The data range from 02/1983 until 12/2002, yielding 239 returns. The size of the benchmark is denoted by $N$. All numbers are annualized.

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

Table 2: **Mean-Summary Statistics for Excess Returns with Gain = 300 bp**
This table presents *ex post* information ratios, means, and standard deviations of realized excess returns. The gain (i.e., the expected excess return) was fixed at 300 basis points. The out-of-sample period is 02/1983 until 12/2002, yielding 239 monthly excess returns. The size of the benchmark is denoted by $N$. 'Sample' denotes the sample covariance matrix; 'Shrink-CC' denotes our shrinkage estimator (2); 'Shrink-SF' denotes the shrinkage estimator of Ledoit and Wolf (2003); 'PC-5' denotes the estimator based on the first five principal components. The results are mean-summaries over 50 repetitions. All numbers are annualized.

|  | 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 |

Table 3: **Mean-Summary Statistics for Average Monthly Turnover** This table presents average monthly turnovers for various strategies. The gain (i.e., the expected excess return) was fixed at 300 basis points. The out-of-sample period is 02/1983 until 12/2002, yielding 239 monthly portfolio updates. The size of the benchmark is denoted by $N$. 'Sample' denotes the sample covariance matrix; 'Shrink-CC' denotes our shrinkage estimator (2); 'Shrink-SF' denotes the shrinkage estimator of Ledoit and Wolf (2003); 'PC-5' denotes the estimator based on the first five principal components. The results are mean-summaries over 50 repetitions.

|  | $N = 30$ | $N = 50$ | $N = 100$ | $N = 225$ | $N = 500$ |
|---|---|---|---|---|---|
| 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 |

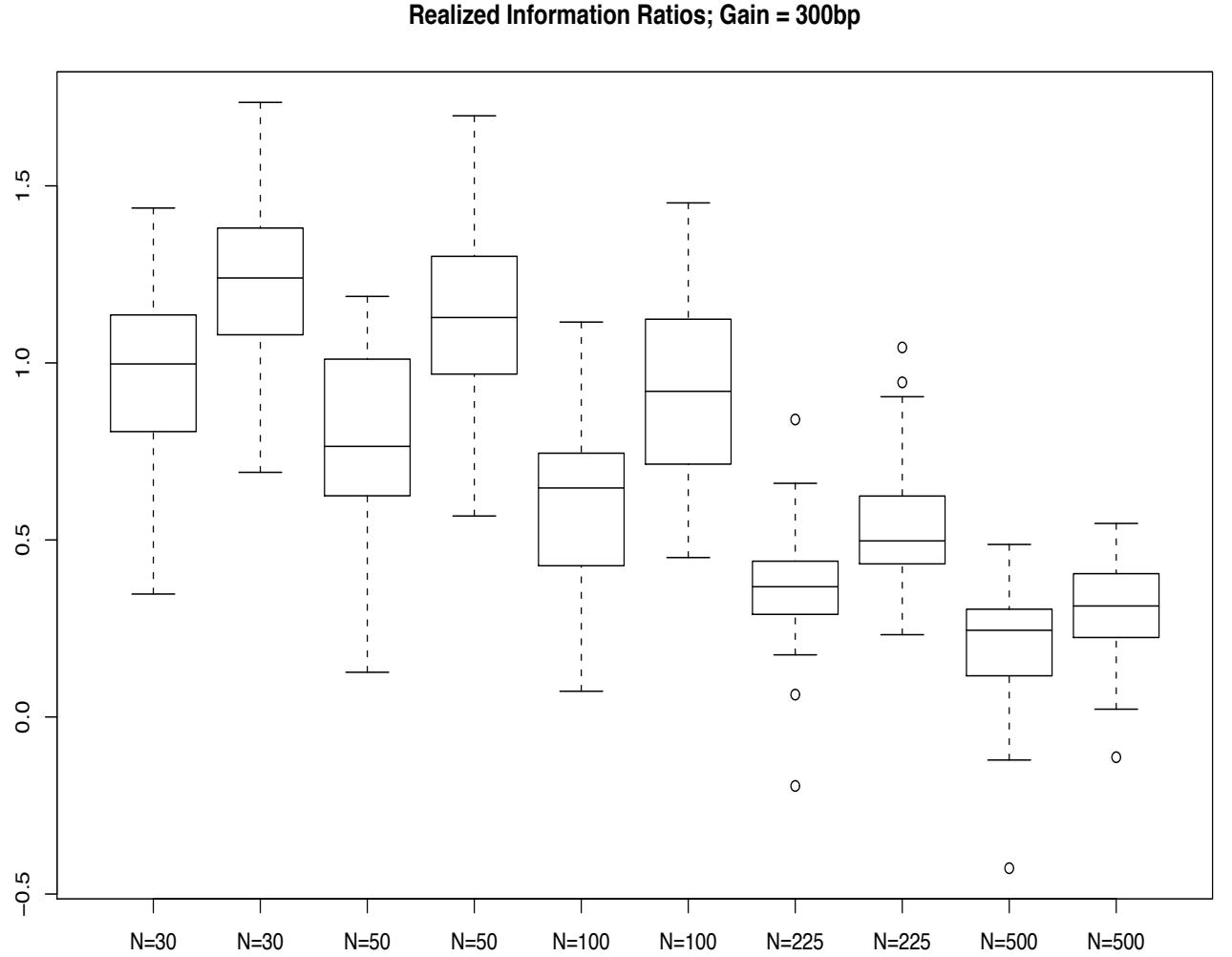**Realized Information Ratios; Gain = 300bp**



Figure 1: Boxplots of the realized information ratios with a gain of 300 basis points. For any given index size $N$, the first boxplot corresponds to the sample covariance matrix and the second one corresponds to the shrinkage estimator (2). The plots show the 50 repetitions which gave rise to the mean-summaries in Table 2.