

Shrinking the Covariance Matrix—Simpler is Better

Final version published in
**Journal of Portfolio
Management**, Summer
2007, Vol 33, No. 4, pp.
56-63.

David J. Disatnik
Finance Department
Faculty of Management
Tel Aviv University, Israel
tel: +972-546-566-185
daveydis@post.tau.ac.il

and

Simon Benninga
Professor of Finance
Faculty of Management
Tel Aviv University, Israel
tel: +972-523-384-690
benninga@post.tau.ac.il

This version: 27 October 2006

Abstract

This paper deals with the construction of the covariance matrix for portfolio optimization. We show that in terms of the ex-post standard deviation of the global minimum variance portfolio, there is no statistically significant gain from using more sophisticated shrinkage estimators instead of simpler portfolios of estimators. This is true both when short sale constraints that prevent the portfolio weights from being negative are imposed as well as when they are not imposed.

We thank Michael Brandt, Dessi Pachamanova (our referee), Michael Wolf and seminar participants at the University of Basel, the Hebrew University of Jerusalem, the University of Konstanz and Tel Aviv University for helpful comments. We also thank the Kurt Lion Foundation for research support.

Shrinking the Covariance Matrix—Simpler is Better

Background

The computational aspects of finding efficient portfolios have been a concern of the finance profession since the seminal work of Markowitz [1952, 1959]. While the mathematics of efficient portfolios are relatively simple, the traditional practical implementation of the theory often leads to questionable results—in many cases the covariance matrix is not invertible, and in other cases the "optimal" portfolios have very large short sale positions.

Essentially there are two approaches to deal with the problematic implementation results. The first is the "theoretical approach," in which theoretical aspects and assumptions regarding portfolio optimization are re-examined. The second approach is the "implementation approach," which largely stems from the fact that the two main elements of Markowitz's mean-variance (MV) theory—the expected stock returns vector and the covariance matrix of the stock returns—are unknown, and thus must be estimated. This paper is part of this second, implementationally-oriented literature, and it mainly deals with the estimation of the covariance matrix.

The estimation of the covariance matrix, like any other estimation process, contains an error. When discussing this error, it is common to distinguish between estimation error and specification error. The estimation error occurs when there are not enough degrees of freedom per estimated parameter, or in other words when the number of observations in the sample is not big enough compared to the number of the estimated parameters. The specification error occurs, when some form of structure is imposed on the model that is being used in the estimation process, and therefore the estimator becomes too specific in comparison with reality.

The traditional and probably the most intuitive estimator of the covariance matrix is the sample covariance (henceforth—the sample matrix). However, as Pafka et al. [2004] state, this estimator often suffers from the "curse of dimensions": In many cases the length of the stock returns' time series used for estimation (T) is not big enough compared to the number of stocks considered (N). As a result, the estimated covariance matrix is ill conditioned. Michaud [1989] points out that inverting such a matrix (as required by the MV theory) amplifies the estimation error tremendously. Furthermore, when N is bigger than T , the sample covariance matrix is not even invertible at all.¹

The literature dealing with methods to improve the estimation of the covariance matrix is too extensive to survey here. In recent years several studies have concentrated on estimators using monthly data under the assumptions of return stationarity and that sample variances are good estimators of the stock variances. Basically, most of these studies stem from a fundamental principle of statistical theory—there exists a tradeoff between the estimation error and the specification error. Hence, in order to develop an improved estimator, the huge estimation error of the sample matrix must be reduced without creating too much specification error. Combining the findings of the studies of Chan et al. [1999], Bengtsson and Holst [2002], Jagannathan and Ma [2003], Ledoit and Wolf [2003], Ledoit and Wolf [2004] and Wolf [2004] suggests that the best estimators of this type are shrinkage estimators and portfolios of estimators.

The roots of the shrinkage method in statistics are not related to covariance estimation and can be found in the seminal work of Stein [1955].² Roughly speaking, in our context, a shrinkage estimator is usually a weighted average of the sample matrix with an invertible

¹ See for example Ledoit and Wolf [2003].

² Efron and Morris [1977] give a beautiful description of Stein's work.

covariance matrix estimator on which quite a lot of structure is imposed and whose diagonal elements are the sample variances.³ The proportions used in the shrinkage estimator are often found by minimizing the quadratic risk (of error) function of the combined estimator. These proportions are supposed to guarantee the reduction of the estimation error of the sample matrix without creating instead too much specification error (which is related to the second estimator in the weighted average). As a result the off-diagonal elements of the shrinkage estimator are moderated (or shrunk) compared to the typically large off-diagonal elements of the sample matrix. The variance elements in the diagonal are kept untouched.

Jagannathan and Ma [2000] use the concept of a portfolio of covariance estimators. A portfolio of estimators is an estimator consisting of an equally weighted average of the sample matrix and several other estimators of the covariance matrix whose diagonal elements are the sample variances and at least one of them is invertible. This concept has also been adopted by Bengtsson and Holst [2002]; it is based on the logic that estimators based on different assumptions make errors in different directions. The portfolio of estimators diversifies among these errors, and they hopefully cancel out. In essence, the portfolio approach builds on the tradeoff between estimation and specification error. By averaging the sample matrix (which suffers from much estimation error) with other estimators whose primary error is specification error an improved covariance matrix can be obtained.

Both the shrinkage estimators and the portfolios of estimators which appear in the literature have been shown to perform substantially better than the sample matrix. However, the shrinkage estimators are more complex than the portfolios of estimators, at least in their

³ It is straightforward to show that a weighted average of two matrices, one of which is invertible, is also invertible. Thus, the shrinkage estimator is always invertible.

theoretical derivation. While a portfolio of estimators is often simply derived by using an equally weighted average, the derivation of a shrinkage estimator involves solving a minimum problem for finding the proportions in the weighted average and estimating these proportions, as they depend on some unknown parameters.⁴

In this paper we check whether, in terms of performance, there is any gain from using the more sophisticated shrinkage methods. We do this by running a performance contest, which is based on a "horse race" between several shrinkage estimators and portfolios of estimators. We use the ex-post standard deviation of the global minimum variance portfolio (GMVP) as our betterment criterion. We show that all the estimators perform within the same range, and that it is actually impossible to claim that one of them is the better than the other. Our conclusion is that one can use the simpler estimators rather than the more complicated estimators. That is, there is no real need to use the shrinkage estimators and instead one can simply use the portfolios of estimators.

A significant drawback of many covariance matrix estimators, including the shrinkage estimators and the portfolios of estimators, is that they generate minimum variance portfolios incorporating significant short sale positions. Short selling is a significant implemental problem in portfolio computations: It is widely prohibited (mutual funds, for example, in many cases are not allowed to short sell) and many individual investors find short selling onerous or impossible. Therefore, to the extent that short sales are indeed considered an undesirable feature of portfolio optimization, there is some interest in finding an estimator of the covariance matrix that performs substantially better than the sample matrix and produces positive efficient portfolios.

⁴ An example of the proportions estimator can be found Ledoit and Wolf [2003]. While the shrinkage estimators may appear to be computationally complex, these are easily computed using Matlab.

Probably the most intuitive way to obtain positive portfolios is to add short sale constraints to the portfolio selection problem.⁵ In order to check empirically whether this way can also produce estimators that generate relatively low ex-post standard deviations of the GMVP, we run a new contest—this time imposing the short sale constraints, and using the following estimators: the sample matrix, one shrinkage estimator, one portfolio of estimators and a matrix containing only the diagonal elements of the sample matrix (henceforth—the diagonal matrix), which serves as our "stalking horse," since it contains much specification error.⁶

Similarly to results reported in Bengtsson and Holst [2002] and Jagannathan and Ma [2003], we find that when the short sale constraints are imposed, all three estimators perform substantially better than the diagonal matrix. Not surprisingly, however, imposing short sale constraints has a cost: When compared to the unconstrained GMVP constructed from the shrinkage estimators and the portfolios of estimators, the GMVP has significantly higher variance in the presence of the short sale constraints. This is true no matter which of the three estimators we use when the short sale constraints are imposed. This statistically significant gap

⁵ Jagannathan and Ma [2003] show that imposing such constraints can be interpreted as a means of shrinking. They show that constructing the constrained GMVP from the sample matrix is equivalent to constructing the unconstrained GMVP from a shrunk covariance matrix.

⁶ Note that when the short sale constraints are imposed, the numerical solution for the weights of the GMVP does not depend on inverting the covariance matrix. In this case, therefore, the sample matrix can be added to the performance contest no matter the number of stocks considered.

between the performances of the estimators when the short sale constraints are imposed and not imposed is the "price" of not holding short sale positions.⁷

Our findings differ from those of Bengtsson and Holst [2002] and Jagannathan and Ma [2003] in at least one significant respect: When the short sale constraints are imposed, both the shrinkage estimator and the portfolio of estimators perform statistically significantly better than the sample matrix. We also find that, when imposing the constraints, the portfolio of estimators performs at least as well as the more sophisticated shrinkage estimator. This again confirms our notion that simpler is better, at least when it comes to shrinkage.

The remainder of this paper proceeds as follows: First, we describe our data, our methodology, and the covariance matrix estimators used in our study. Then, we present the results of our "horse race" between the shrinkage estimators and the portfolios of estimators with and without short sale constraints, respectively. We conclude the paper with a brief summary.

Data and period of study

We use monthly returns on stocks traded on the New York Stock Exchange (NYSE). The stock returns are extracted from the Center for Research in Security Prices (CRSP) database. The period of the study is from 1/1964 to 12/2003.

⁷ Levy and Ritov [2001] also discuss the "price" of not holding short positions. They measure this price by comparing the Sharpe ratios of the optimal portfolios when the short sale constraints are imposed and not imposed.

Methodology

For evaluating the performance of the estimators included in our performance contest, one can compare the estimators computed over a certain sample period (the in-sample period) with the covariance matrix realized over a subsequent period (the out-of-sample period). However, our main interest is to assess how the performance of the estimators translates into the performance of the optimal portfolios obtained from the MV optimization process. Therefore, we find it more useful to conduct an empirical performance contest that focuses on the ex-post performance of the respective optimal portfolios. We follow Chan et al. [1999], Bengtsson and Holst [(2002], Jagannathan and Ma [2003], and Ledoit and Wolf [2003], and use the ex-post standard deviation of the GMVP as our betterment criterion. We benefit from the fact that finding the GMVP does not require the estimation of the expected stock returns vector, which is out of the scope of our paper.⁸

In our contest we mimic an investor who invests in the GMVP of stocks traded on the NYSE. Our investor chooses a covariance matrix estimator which is based on historical data of stock returns and then chooses the length of the in-sample period, the historical time frame over which he collects monthly return data. Based on this data our investor computes the covariance matrix estimator, finds the GMVP, and invests in this portfolio.

Our investor has a fixed investment horizon during which he keeps his portfolio unchanged (the out-of-sample period). When this period is over, he liquidates the portfolio and

⁸ Elton et al. [2005] point out that evaluating the performance of the covariance matrix estimators based on the ex-post GMVP suffers from some bias, since the stocks enter the GMVP are principally the ones with low correlations. Basak et al. [2004] state that the estimate of the variance of a GMVP constructed using an estimated covariance matrix will on average be strictly smaller than its true variance.

starts the whole process of estimating the covariance matrix, constructing the GMVP and holding it until liquidation all over again.

To illustrate the way we conduct our performance contest, let us assume that the first time our investor wishes to invest is January 1974, and that he chooses an in-sample period of 120 months and an out-of-sample period of 12 months. Then:

1. We collect monthly return data of stocks traded on the NYSE from 1/64 till 12/73. We choose an estimator of the covariance matrix, which is computed based on this data.
2. We construct the GMVP from the estimator computed in phase 1.
3. We record the monthly returns on the GMVP From 1/74 till 12/74.
4. We start the whole process all over again. Namely, we collect monthly return data of stocks traded on the NYSE from 1/65 till 12/74, based on this data we compute the same estimator used in phase 1, construct the GMVP and record its monthly returns from 1/75 till 12/75 and so on.
5. We repeat the process of computing the covariance matrix, constructing the GMVP and recording its monthly returns in the out-of-sample period 30 times (the last monthly return recorded is from 12/2003). As a result, all together, we collect 360 monthly returns on the GMVP (from 1/74 till 12/2003).
6. We compute the standard deviation of the collected 360 monthly returns. This ex-post standard deviation represents the risk our investor was exposed to in the 30 years he was running his investment strategy. Given the chosen in-sample and out-of-sample periods, we can refer to the computed ex-post standard deviation as a proxy of the performance of the specific estimator used for estimating the covariance matrix.

We conduct our test for several estimators. Since the motivation of our investor is to minimize the risk of his investment, the smaller the standard deviation of the collected 360 monthly returns, the better the respective estimator of the covariance matrix.⁹

We run our "horse race" six times, each time changing the length of the in-sample period or the length of the out-of-sample period. We use in-sample periods of 120 months (also used in Ledoit and Wolf [2003]) and 60 months (also used in Chan et al. [1999] and Jagannathan and Ma [2003]).¹⁰ We use out-of-sample periods of 12 months (also used in Chan et al. [1999], Jagannathan and Ma [2003] and Ledoit and Wolf [2003]), 24 months and 36 months. We chose these three out-of-sample periods, since we believe they correspond to realistic investment horizons (see also Chan et al. [1999]). As an aside, we always construct the first GMVP on 1/74 and record the last return data on the last GMVP on 12/03. Thus, in each of the six runs for every one of the seven estimators participating in our contest, we have a set of 360 monthly returns used to compute the respective ex-post standard deviation.

It is also worth mentioning that each time we construct a GMVP, we construct it only out of NYSE stocks whose returns cover the entire in-sample and out-of-sample periods used. For example, in the case of in-sample period of 120 months and out-of sample period of 12 months, for constructing the GMVP of 1/74, we only use NYSE stocks with monthly return data for all

⁹ It might appear more intuitive to calculate the covariance matrix realized in the out-of-sample period. Then finding the GMVP generated from the calculated covariance matrix and checking which covariance matrix estimator generated the "closest" GMVP to the calculated GMVP. However, this type of test cannot be done when the number of stocks considered is bigger than the stock returns' time series used (as in our case), since then the realized covariance matrix is not invertible and the GMVP cannot be calculated.

¹⁰ Jobson and Korkie [1981] mention rules of thumb regarding the length of the in-sample period of 4 to 7 years and 8 to 10 years.

the 132 months from 1/64 till 12/74. For constructing the GMVP of 1/75, we only use NYSE stocks with monthly return data for all the 132 months from 1/65 till 12/75 and so on. Therefore, the resulting number of stocks used for constructing the GMVP varies across the years and the runs of the contest (see also Bengtsson and Holst [2002]).¹¹

The covariance matrix estimators included in our study

In our study we use the following seven covariance matrix estimators:

Shrinkage to the single-index model: This is the shrinkage estimator suggested by Ledoit and Wolf [2003], in which the covariance matrix estimator obtained from Sharpe's [1963] single-index model (henceforth—the single-index matrix) joins the sample matrix in the weighted average.¹² This estimator performed best in Ledoit and Wolf's [2003] contest and best in Jagannathan and Ma's [2003] contest.¹³

Shrinkage to the constant correlation model: This is the shrinkage estimator suggested by Ledoit and Wolf [2004]), in which the covariance matrix estimator obtained by assuming that each pair of stocks has the same correlation (henceforth—the constant correlation matrix) joins

¹¹ We are aware of the fact that this widely-followed procedure introduces survivorship bias into the estimation procedure. However, since the survivorship bias is common to all the compared estimators, we do not consider this a significant problem.

¹² We use the value-weighted portfolio of stocks included in our study as the index.

¹³ In fact, in Jagannathan and Ma's [2003] contest, the shrinkage estimator performed best together with the sample covariance matrix based on daily return data, which is out of the scope of this paper.

the sample matrix in the weighted average. Ledoit and Wolf [2004] find the performance of this estimator comparable to the performance of the shrinkage to the single-index model estimator.

A portfolio of the sample matrix, the single-index matrix and the diagonal matrix: This is the estimator, which had been introduced by Jagannathan and Ma [2000] and adopted by Bengtsson and Holst [(2002]. It consists of an equally weighted average of the sample matrix, the single-index matrix and the diagonal matrix. Its performance was found to be one of the best in Bengtsson and Holst's [2002] contest.

A portfolio of the sample matrix, the single-index matrix and the constant correlation matrix: This estimator has not been previously been used in the literature. It consists of an equally weighted average of the sample matrix, the single-index matrix and the constant correlation matrix.

A portfolio of the sample matrix, the single-index matrix, the constant correlation matrix and the diagonal matrix: This estimator has not been introduced in the literature before. It consists of an equally weighted average of the sample matrix, the single-index matrix, the constant correlation matrix and the diagonal matrix.

A random average of the sample matrix and the single-index matrix: As in the case of the shrinkage to the single-index model estimator, this estimator is based on a weighted average of the sample matrix and the single-index matrix. However, this time the proportion of the single-index matrix in the weighted average is drawn from a uniform distribution on the interval (0.5,1). We include this estimator in our contest, because it helps us highlight the problem of estimating the proportions of the estimators in the shrinkage estimators (see also Bengtsson and Holst [2002] and Jagannathan and Ma [2003]). Ledoit and Wolf [2003] and Bengtsson and Holst [2002] point out that there is more estimation error in the sample matrix than there is

specification error in the single-index matrix. Therefore, we allow the proportion of the single-index matrix in the random average to obtain only values greater than 0.5.

The diagonal matrix: This estimator contains a lot of specification error, and clearly does not obey the statistical principle of reducing the estimation error of the sample matrix without creating instead too much specification error. It is included in the contest as our "stalking horse," since for our data the sample matrix is not invertible and cannot therefore be used.

The performance contest with no short sale constraints

In this section we describe the results of our "horse race", when the short sale constraints are not imposed. We show empirically that there is no statistically significant gain from using the more sophisticated shrinkage methods—all of the methods discussed in this section lead to similar improvements in terms of the ex-post standard deviation of the GMVP. We conclude that simpler is better, at least when it comes to shrinkage.

In Exhibit 1 we report the ex-post standard deviations obtained for each covariance matrix estimator, when the short sale constraints are not imposed. The standard deviations are annualized through multiplication by $\sqrt{12}$.

Ex-Post Standard Deviations, Short Sales Constraints Are Not Imposed						
	In sample 120 months			In sample 60 months		
	Out of sample period					
	12 months	24 months	36 months	12 months	24 months	36 months
Average size of universe of stocks	901	853	806	1261	1186	1110
The largest universe	1063	945	865	1739	1572	1424
The smallest universe	795	769	729	1029	985	958
Diagonal	13.12%	13.12%	13.16%	12.90%	12.92%	12.84%
Shrinkage to constant correlation	8.52%	8.97%	8.91%	8.46%	8.85%	8.83%
Random average of sample and single index	8.51%	8.93%	9.00%	8.47%	8.83%	8.89%
Portfolio of sample, single index, constant correlation	8.47%	8.90%	8.85%	8.40%	8.83%	8.81%
Portfolio of sample, single index, constant correlation, diagonal	8.46%	8.85%	8.81%	8.37%	8.81%	8.78%
Portfolio of sample, single index, diagonal	8.39%	8.89%	8.93%	8.34%	8.94%	8.91%
Shrinkage to single index	8.37%	8.89%	8.94%	8.31%	8.91%	8.90%
Gap between worst and best improver	0.15%	0.12%	0.19%	0.16%	0.13%	0.13%

Exhibit 1: The annualized ex-post standard deviations generated by each of the seven tested covariance matrix estimators in the six runs of the performance contest, when the short sale constraints are not imposed.

The six estimators all perform substantially better than the "stalking horse," the diagonal matrix. The most important point of our comparison, however, is not the improvement of the six methods over the (obviously inferior) diagonal matrix. The most important point is that all of the estimators we examine perform within the same range. In other words, we find very little qualitative difference between any of the estimators: The largest gap in a specific run between the standard deviations corresponding to the "best" and "worst" improvements of our six estimators is obtained in the case of in-sample period of 120 months and out-of-sample period of 36 months and it is only 0.19%. Checking, for each run, whether the tiny differences in the performance of the various estimators are statistically significant results in a negative answer in all cases.¹⁴ In addition, we can see that the ranking of the estimators changes from one run to another, and in fact the portfolio of the sample matrix, the single-index matrix, the constant

¹⁴ Throughout this paper, we use chi square tests for statistical inference. The smallest p-value obtained here is 29.12%.

correlation matrix and the diagonal matrix "wins" the contest four times, whereas the shrinkage to the single-index model estimator "wins" only twice.

We conclude that all six of the estimators have substantially the same performance improvement, and that therefore there is no need to use the methodologically complex shrinkage estimators. Instead, one can simply use the portfolio of estimators, as long as these portfolios obey the statistical principle of reducing the estimation error of the sample matrix without creating instead too much specification error.

One could claim that if a better estimator than the market matrix or the constant correlation matrix is found, then the shrinkage estimator relying on this estimator will perform better than any portfolio of estimators. We do not agree with such a claim. In our opinion, placing this estimator in a portfolio of estimators will result in a performance within the same range as of the shrinkage estimator. An example for that can be found in the work of Bengtsson and Holst [2002]. They develop a rather complicated shrinkage estimator, in which the estimator joining the sample matrix is generated from principal component analysis. According to them, their new shrinkage estimator performs better than the shrinkage estimator of Ledoit and Wolf [2003]. However, when they use a portfolio of estimators consisting of the estimator generated from principal component analysis, the sample matrix and the diagonal matrix, they obtain an estimator that performs at least as good as their shrinkage estimator.

We can see that both the random average of the sample matrix and the single-index matrix and the shrinkage to the single-index model estimator perform within the same range. Theoretically, the shrinkage estimator should perform better than any other weighted average of the two estimators, since the proportions in the weighted average of the shrinkage estimator are obtained from minimizing the quadratic risk (of error) function of the combined estimator. Yet,

it seems that in practice, estimating these specific proportions gives rise to a new type of error, and overall the shrinkage estimator does not perform better than the random average. This result is similar to the one obtained regarding this issue in Jagannathan and Ma [2003]. We suspect Bengtsson and Holst (2002) report a contradicted result, because they use for their random average a uniform distribution on the interval $(0,1)$, and not on the interval $(0.5,1)$. Thus, they do not prevent situations, in which the proportion of the single-index matrix is smaller than the proportion of the sample matrix. In these cases, the estimator obtained contains probably too much estimation error, and therefore cannot compete with the shrinkage estimator.

As an aside, we can notice that when keeping the length of the out-of-sample period fixed and changing the in-sample period from 120 to 60 months, we obtain quite similar results for the performance of the various estimators. When keeping the in-sample period fixed and changing the out-of-sample period from 12 months to 24 or 36 months, the annualized standard deviations grow in approximately 0.45%, and the p-values for the chi-square tests for significance of differences in performance range from 2.02% to 12.71%. We believe future research should address more carefully the effects (if any) of the chosen length of the in-sample and out-of-sample periods on the performance of the covariance matrix estimators.

The performance contest with short sale constraints

So far we have evaluated the covariance matrix estimators without addressing the issue of short selling positions. In this section we discuss this issue and the effects of imposing the short sale constraints on the portfolio selection problem.

Average short positions

In Exhibit 2 we present the average amount of short positions obtained for each estimator in all six runs of our contest. The amount of short positions is defined as the sum of all negative portfolio weights.

Average Short Positions						
<i>Covariance Matrix Estimator</i>	<i>In Sample / Out of Sample</i>					
	120 / 12	120 / 24	120 / 36	60 / 12	60 / 24	60 / 36
Diagonal	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Random average of sample and single index	-90.09%	-91.40%	-87.84%	-62.01%	-63.37%	-64.21%
Portfolio of sample, single index, constant correlation	-117.17%	-117.46%	-116.50%	-91.36%	-92.10%	-94.46%
Portfolio of sample, single index, diagonal	-90.21%	-90.47%	-89.64%	-65.22%	-65.86%	-67.87%
Portfolio of sample, single index, constant correlation, diagonal	-101.41%	-101.23%	-99.99%	-84.04%	-84.43%	-86.30%
Shrinkage to constant correlation	-129.07%	-130.24%	-130.94%	-92.76%	-93.41%	-96.44%
Shrinkage to single index	-96.96%	-97.96%	-97.84%	-66.55%	-67.19%	-69.29%

Exhibit 2: Average short positions—total of all negative portfolio proportions—generated by each of our seven tested covariance matrix estimators. An average short interest of -96.96%, obtained for the shrinkage to the single-index model estimator in the run of in-sample period of 120 months and out-of-sample period of 12 months, means that in average, over the 30 portfolios constructed in this run based on this estimator, for every dollar invested in the portfolio we short 96.96 cents worth of stocks, while buying \$1.9696 worth of other stocks.

We can see that, apart from the estimator based on the diagonal matrix, which generates a positive GMVP, all other estimators in all the runs of the contest generate portfolios with significant short sale positions. Moving from in-sample period of 120 months to in-sample period of 60 months reduces the average short positions for each one of the estimators; however even then we are still talking about quite significant short sale positions. Also Bengtsson and Holst [2002], Jagannathan and Ma [2003] and Ledoit and Wolf [2003] report significant short positions in their performance contests.

To the extent that short sales are considered an undesirable feature of portfolio optimization, the shrinkage estimators and the portfolios of estimators cannot satisfy us anymore.

We cannot count on the diagonal estimator either, since it generates relatively high out-of sample standard deviations (see Exhibit 1). Hence, our goal now is to check whether we can find an estimation method that generates both low ex-post standard deviations and positive portfolios. One candidate for such a method is the addition of the short sale constraints to the GMVP problem. In the next subsection we examine empirically the attractiveness of this method.

Imposing the short sale constraints

We again run our performance contest, this time imposing the short sale constraints. We focus this time on three covariance matrix estimators—the sample matrix, the shrinkage to the single-index model estimator and the portfolio of the sample matrix, the single-index matrix, the constant correlation matrix and the diagonal matrix.¹⁵ As before, our "stalking horse" is the diagonal matrix and the ex- post standard deviation of the GMVP is used as our betterment criterion.

In Exhibit 3 we report the ex-post standard deviations obtained for each covariance matrix estimator, when the short sale constraints are imposed. The standard deviations are annualized through multiplication by $\sqrt{12}$.

¹⁵ In order to find the GMVP weights, we use the Mosek iterative procedure together with the Matlab program. When the diagonal matrix is used, the short sale constraints are of course not needed, as the diagonal matrix anyhow generates a positive GMVP.

Ex-Post Standard Deviations, Short Sales Constraints Are Imposed						
	In sample 120 months			In sample 60 months		
	Out of sample period					
	12 months	24 months	36 months	12 months	24 months	36 months
Average size of universe of stocks	901	853	806	1261	1186	1110
The largest universe	1063	945	865	1739	1572	1424
The smallest universe	795	769	729	1029	985	958
Diagonal	13.12%	13.12%	13.16%	12.90%	12.92%	12.84%
Sample	10.74%	10.87%	11.08%	10.84%	11.23%	11.48%
Shrinkage to single index	9.94%	10.19%	10.23%	9.75%	10.24%	10.17%
Portfolio of sample, single index, constant correlation, diagonal	9.65%	10.01%	10.06%	9.15%	9.65%	9.57%

Exhibit 3: The annualized ex-post standard deviations generated by each of the four tested covariance matrix estimators in the six runs of the performance contest, when the short sale constraints are imposed.

As in the studies of Bengtsson and Holst [2002] and Jagannathan and Ma [2003], our results confirm that imposing the short sale constraints substantially reduces the ex-post standard deviations compared to the GMVP of our "stalking horse," the diagonal matrix. Not surprisingly, however, imposing short sale constraints has a cost, which is revealed when one compares the ex-post standard deviations generated by the shrinkage estimator and the portfolio of estimators when the short sale constraints are imposed and when they are not imposed. In the case of the shrinkage estimator, imposing the constraints increases the ex-post standard deviations by about 1.3% to 1.6%. In the case of the portfolio of estimators, imposing the constraints increases the ex-post standard deviations by about 1.2% when in-sample periods of 120 months are used and by about 0.8% when in-sample periods of 60 months are used. All gaps in all runs for both estimators are statistically significant.¹⁶ This statistically significant gap between the performances of the estimators when the short sale constraints are imposed and not imposed is the "price" of not holding short sale positions.

¹⁶ We again use a chi square test. The largest p-value obtained this time is 0.39%.

Our findings differ from those of Bengtsson and Holst [2002] and Jagannathan and Ma [2003] in at least one significant respect: When the short sale constraints are imposed, both the shrinkage estimator and the portfolio of estimators perform statistically significantly better than the sample matrix.¹⁷ In contrast, applying the same statistical significance tests to the results and samples of Bengtsson and Holst [2002] and Jagannathan and Ma [2003] reveals that in these studies the gaps obtained are not statistically significant. We believe future research should address this issue more carefully. We can also see that systematically the portfolio of estimators generates lower ex-post standard deviations than the shrinkage estimator, with p-values that range from 4.18% to 31%. This again confirms our notion that simpler is better, at least when it comes to shrinkage.

As an aside, it can be noticed that when keeping the out-of-sample period unchanged and reducing the in-sample period from 120 months to 60 months, the portfolio of estimators performs better, the performance of the shrinkage to the single-index model estimator is almost unchanged and the sample matrix deteriorates. This time the smallest p-value obtained is 6.12%. In addition, keeping the in-sample period unchanged and increasing the out-of-sample period from 12 months to 24 or 36 months damage the performance of the three estimators in terms of the ex-post standard deviations. The smallest p-value obtained for this set of checks is also 6.12%.

¹⁷ In the run of in-sample period of 120 months and out-of-sample period of 24 months, for the gap between the shrinkage estimator and the sample matrix we obtain a p-value of 2.87%. In all the other cases, the largest p-value obtained is 1.07%.

Summary

This paper deals with estimating the covariance matrix of stock returns, which is one of the two main elements of the mean-variance theory of Markowitz [1952, 1959] for portfolio selection.

Estimating the covariance matrix based solely on the sample matrix is famously difficult, since very often the sample matrix suffers from the "curse of dimensions." As a result, a significant finance literature, which looks for better methods to estimate the covariance matrix, has been spawned. Out of this rich literature, we have chosen to focus, in this paper, on estimators using monthly data under the assumptions of return stationarity and that sample variances are good estimators of the stock variances. Combining the findings of recent studies reveals that the best estimators of that type are the shrinkage estimators and the portfolios of estimators.

In our study, we run a "horse race" between various shrinkage estimators and portfolios of estimators. We use the ex-post standard deviation of the global minimum variance portfolio (GMVP) as our betterment criterion. We show empirically that all the estimators perform within the same range, and that it is actually impossible to claim that one of them is the better than the other. Hence, there is no statistically significant gain from using the more sophisticated shrinkage methods, and therefore one can instead use the simpler portfolios of estimators. We conclude that simpler is better, at least when it comes to shrinkage.

A significant drawback of many covariance matrix estimators, including the shrinkage estimators and the portfolios of estimators, is that they generate minimum variance portfolios incorporating significant short sale positions. To the extent that short sales are considered an undesirable feature of portfolio optimization, the most intuitive way to overcome them is to add

to the portfolio selection problem short sale constraints that prevent the portfolio weights from being negative, no matter which covariance matrix estimator is used.

In our study, we also run a "horse race" in which the short sale constraints are imposed and again the ex-post standard deviation of the GMVP is used as the betterment criterion. Our findings regarding the short sale constraints differ from previous studies in at least one significant respect: In our sample when the short sale constraints are imposed, both the shrinkage estimator and the portfolio of estimators perform statistically significantly better than the sample matrix. We believe future research should address this issue more carefully. We also find that, when imposing the constraints, the portfolio of estimators performs, ex-post, at least as well as the more sophisticated shrinkage estimator. This again confirms our notion that simpler is better, at least when it comes to shrinking the covariance matrix.

References

- Basak, G.K., R. Jagannathan and T. Ma. [2004]. Assessing the Risk in sample Minimum Risk Portfolios. *Working paper*, University of Bristol, Northwestern University and University of Utah.
- Bengtsson, C. and J. Holst. [2002]. On portfolio selection: Improved Covariance Matrix Estimation for Swedish Asset Returns. *Working paper*, Lund University and Lund Institute of Technology.
- Chan, L. K. C., J. Karceski and J. Lakonishok. [1999]. On Portfolio Optimization: Forecasting Covariances and Choosing the Risk Model. *Review of Financial Studies*, 12: 937-974.
- Efron, B. and C. Morris. [1977]. Stein's Paradox in Statistics. *Scientific American*, 236: 119-128.
- Elton, E. J., M. J. Gruber and J. Spitzer. [2005]. Improved Estimates of Correlation Coefficients and Their Impact on the Optimum Portfolios. *Working paper*, New York University.
- Jagannathan, R. and T. Ma. [2000]. Three Methods for Improving the Precision in Covariance Matrix Estimation. *Unpublished working paper*.
- Jagannathan, R. and T. Ma. [2003]. Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps. *Journal of Finance*, 54(4): 1651-1683.
- Jobson, J. D. and B. Korkie. [1981]. Putting Markovitz theory to work. *Journal of Portfolio Management*, 7: 70-74.
- Ledoit, O. and M. Wolf. [2003]. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, 10: 603-621.
- Ledoit, O. and M. Wolf. [2004]. Honey, I Shrunk the Sample Covariance Matrix. *Journal of Portfolio Management*, 30(4): 110-119.

- Levy, M. and Y. Ritov. [2001]. Portfolio optimization with many assets: the importance of short-selling. *Working Paper*, the Hebrew University of Jerusalem.
- Markowitz, H. M. [1952]. Portfolio Selection. *Journal of Finance*, 7: 77-91.
- Markowitz, H. M. [1959]. *Portfolio Selection: Efficient Diversification of Investments*. Yale University Press, New Haven, CT.
- Michaud, R. O. [1989]. The Markowitz Optimization Enigma: Is “Optimized” Optimal? *Financial Analysts Journal*, 45: 31-42.
- Pafka, S., M. Potters, and I. Kondor. [2004]. Exponential Weighting and Random-Matrix-Theory-Based Filtering of Financial Covariance Matrices for Portfolio Optimization. <http://arxiv.org/abs/cond-mat/0402573>.
- Sharpe, W. [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. *Proceedings of the 3rd Berkley Symposium on Probability and Statistics*, Berkeley: University of California Press, 197-206.
- Wolf, M. [2004]. Resampling vs. Shrinkage for Benchmarked Managers. *Working paper*, Universitat Pompeu Fabra.