COMPARISON OF SEVERAL COVARIANCE MATRIX ESTIMATORS FOR PORTFOLIO OPTIMIZATION

Ka Ki Ng*, Priyanka Agarwal†, Nathan Mullen*, Dzung Du‡, Ilya Pollak*

Purdue University
*School of Electrical and Computer Engineering
†Department of Mathematics
‡School of Aeronautics and Astronautics Engineering and Department of Statistics
West Lafayette, IN 47907, USA

ABSTRACT
Modern portfolio theory dates back to a seminal 1952 paper by H. Markowitz and has been very influential both in academic finance and among practitioners in the financial industry. Given a set of assets, the theory can be used to compute the amount to be invested in each asset in order to construct an optimally diversified portfolio. One of the parameters required in this calculation is the covariance matrix of asset returns which, in any practical application, is unknown and must be estimated from historical data. Due to the fact that financial data is often nonstationary, basing the estimates on historical data over a very long time period may not be advisable. This renders the problem of covariance estimation difficult, especially for large portfolios. A large body of literature exists proposing different covariance estimators. We focus on one frequently cited paper by Ledoit and Wolf [5] which proposes a covariance estimation method and purports improvements over several other methods. We show that this method leads to statistically significant improvements over the following methods. We show that this is not the case: in fact, their method does not exhibit statistically significant differences from three other methods.

Index Terms—Portfolios, Markowitz, statistical significance, stock, market, finance, covariance

1. INTRODUCTION

The gross return of a security over one month is defined as the ratio of the closing prices of the security for that month and for the preceding month. The return of a security is defined as the gross return minus one.

Suppose that we have $N$ risky assets with return vector $\mathbf{R}$, modeled as a random vector with expected return $\mathbf{\mu} = \mathbb{E}[\mathbf{R}]$ and covariance matrix $\mathbf{\Lambda} = \mathbb{E}[(\mathbf{R} - \mathbf{\mu})(\mathbf{R} - \mathbf{\mu})^T]$. In other words, $\mathbf{R} = (R^{(1)}, \ldots, R^{(N)})^T$ where $R^{(n)}$ is the return of the $n$-th asset. We assume that the covariance matrix $\mathbf{\Lambda}$ is invertible. Out of these $N$ assets, we form a portfolio with allocation weights $w = (w^{(1)}, \ldots, w^{(N)})^T$. This means that we invest $w^{(n)}$ into the $n$-th asset. We always assume that the total portfolio weight is one:

$$w^T \mathbf{1} = 1, \quad (1)$$

where $\mathbf{1}$ is an $N$-vector of ones. The total portfolio return is then $\mathbf{w}^T \mathbf{R}$, the expected return is $\mathbf{w}^T \mathbf{\mu}$, and the variance of the return is $\mathbf{w}^T \mathbf{\Lambda} \mathbf{w}$.

The classical Markowitz portfolio framework [7] defines portfolio risk as the standard deviation of the portfolio return, and seeks a portfolio weight vector $\mathbf{w}$ which minimizes the portfolio risk subject to a target expected return $\mu_{\text{tgt}}$:

Find $\mathbf{w}^*$ to minimize $\mathbf{w}^T \mathbf{\Lambda} \mathbf{w}$
subject to $\mathbf{w}^T \mathbf{\mu} = \mu_{\text{tgt}} \quad (2)$

Using Lagrange multipliers to perform minimization (2) subject to the constraints (1) and (3) yields [7, 5, 8]:

$$\mathbf{w}^* = \mathbf{\Lambda}^{-1} \mathbf{m} \mathbf{A}^{-1} \mathbf{c}, \quad (4)$$

where $\mathbf{m} = (\mathbf{\mu}^T, 1^T)$, $A = \mathbf{m}^T \mathbf{\Lambda}^{-1} \mathbf{m}$, and $c = (\mu_{\text{tgt}}^T, 1^T)$.

The global minimum-variance portfolio (GMVP) is obtained by dropping the mean constraint (3) and instead minimizing portfolio risk (2) subject only to the weight normalization constraint (1). This yields the following weight vector:

$$\mathbf{w}_{\text{GMVP}} = \frac{\mathbf{\Lambda}^{-1} \mathbf{1}}{\mathbf{1}^T \mathbf{\Lambda}^{-1} \mathbf{1}} \quad (5)$$

In practice, the expected returns and the covariance matrix of the returns are not known and are therefore estimated from historical data. Many approaches exist for estimating the covariance matrix: using the sample covariance; imposing a parametric model such as a linear factor model with a small number of factors; bootstrapping; and combinations of these methods. To evaluate these approaches, typically the following testing method is adopted:

1: Two parameters are selected, the length of the training window $L$ and the length of the holding window $K$. 
2: Covariance matrix and mean returns\(^1\) are estimated from the historical data for the time period \([t - L, t]\).

3: Portfolio weights are calculated based on the estimated covariance and mean returns, and the resulting portfolio is held for the time period \([t, t + K]\).

4: Steps 2 and 3 are repeated with \(t\) replaced by \(t + K, t + 2K, \ldots, t + pK\).

5: The mean and standard deviation of the portfolio return are estimated based on the entire testing period.

A widely cited publication in this field is [5] which proposed estimating the covariance as a convex combination of the sample covariance and the covariance implied by a single-factor linear model where the only factor is the market return. This method is called *shrinkage to market* in [5]. The evidence to support the claim that this method outperforms several alternatives is presented in Table 1 in [5], reproduced including the caption in our Table 1. This table presents the portfolio standard deviations for the portfolio construction strategy described above: the first column for the GMVP and the second column for a minimum-variance portfolio which uses a mean constraint. The portfolio construction strategies are run over 23 years from August 1972 until July 1995, with holding period \(K = 1\) year. Each time portfolio weights are calculated, \(L = 10\) previous years are used to estimate the covariance matrix and, if needed, mean returns. The universe of stocks used in the simulations is the set of all AMEX and NYSE stocks satisfying the following conditions:

- The monthly stock returns are available through the Center for Research in Security Prices (CRSP) monthly return database [2].
- In order for a stock to be included in the portfolio during the year starting at time \(t\), it must have valid CRSP monthly returns for 10 years preceding \(t\) and a valid Standard Industrial Classification (SIC) code.

The numbers in the table appear to suggest that the shrinkage-to-market method proposed in [5] outperforms all other methods in a statistically significant way. For example, focusing on the bottom two lines, we see that the GMVP risk estimate for the shrinkage-to-market strategy plus twice its standard error \(9.55 + 2 \cdot 0.15 = 9.85\) is smaller than the risk estimate for the next best strategy, shrinkage to identity, minus twice its standard error \(10.21 - 2 \cdot 0.17 = 9.87\).

We show, however, that the standard errors reported in parentheses in Table 1 are incorrect. We use three methods to compute the standard errors for the risk estimates:

- In Section 3.1, we assume that the portfolio risk estimates in Table 1 are correct, and that the portfolio returns are iid Gaussian.

---

\(^1\)Note that for the GMVP construction, the mean returns do not need to be estimated.

### Table 1. Risk of minimum variance portfolios

<table>
<thead>
<tr>
<th></th>
<th>St. dev. unconstrained</th>
<th>St. dev. constrained</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identity</td>
<td>17.75 (0.44)</td>
<td>17.94 (0.42)</td>
</tr>
<tr>
<td>Constant correlation</td>
<td>14.27 (0.19)</td>
<td>16.30 (0.29)</td>
</tr>
<tr>
<td>Pseudoinverse</td>
<td>12.37 (0.23)</td>
<td>13.73 (0.32)</td>
</tr>
<tr>
<td>Market model</td>
<td>12.00 (0.16)</td>
<td>13.77 (0.27)</td>
</tr>
<tr>
<td>Industry factors</td>
<td>10.84 (0.17)</td>
<td>12.32 (0.23)</td>
</tr>
<tr>
<td>Principal components</td>
<td>10.31 (0.16)</td>
<td>11.30 (0.22)</td>
</tr>
<tr>
<td>Shrinkage to identity</td>
<td>10.21 (0.17)</td>
<td>11.11 (0.21)</td>
</tr>
<tr>
<td>Shrinkage to market</td>
<td>9.55 (0.15)</td>
<td>10.43 (0.20)</td>
</tr>
</tbody>
</table>

“Unconstrained” refers to the global minimum variance portfolio, while “constrained” refers to the minimum variance portfolio with 20% expected return. Standard deviation is measured out-of-sample at the monthly frequency, annualized through multiplication by \(\sqrt{12}\) and expressed in percents. Standard errors on these standard deviation estimates are reported in parentheses.

- In Section 3.2, we re-calculate the portfolio risk estimates ourselves, and again assume that the portfolio returns are iid Gaussian.
- In Section 4, we use a bootstrapping method.

These methods produce standard errors that are much larger than those in Table 1. These standard errors render the differences among many of the portfolio risks in Table 1 statistically insignificant. Before presenting our results, we describe in Section 2 each covariance estimator used in Table 1.

### 2. Covariance Estimation Methods: Brief Descriptions

For the *identity* estimator, the estimate of the covariance matrix is a scalar multiple of the identity matrix. Specifically, the mean of the sample variances is used as the multiple in [5].

The *constant correlation* estimator assumes that every pair of stocks in the portfolio has the same correlation coefficient [3]. For a portfolio with \(N\) assets, this model has \(N + 1\) parameters: \(N\) return variances and one correlation.

Both the constrained and unconstrained optimal weights require inverse covariance matrices, as evident from Eqs. (4,5). Since in our experiments the number of stocks \(N\) is much larger than the number of historical returns per stock \(L\), the sample covariance matrix is always singular. The *pseudoinverse* method estimates the inverse covariance matrix as the pseudoinverse of the sample covariance.

The *market model* uses the covariance matrix implied by the following single-factor linear model of stock returns [9]:

\[
r_t^{(n)} = \alpha^{(n)} + \beta^{(n)} r_t^{(mkt)} + u_t^{(n)},
\]

where \(r_t^{(n)}\) is the return for the \(n\)-th stock for month \(t\); \(r_t^{(mkt)}\) is the market return for month \(t\); and \(u_t^{(n)}\) is iid zero-mean
noise independent of the market returns. This yields the following covariance matrix:

$$\lambda^{(mkt)} \beta \beta^T + \Sigma,$$

where $\lambda^{(mkt)}$ is the variance of the market returns, $\beta$ is the vector of all the stocks’ betas, and $\Sigma$ is the diagonal covariance matrix of the noise term in Eq. (6). The variance of the market returns is estimated as the sample variance, the betas are estimated via linear regression of each stock’s returns on the market returns, and the noise variances are estimated as the residual variances in that regression.

The industry factors model uses the covariance matrix implied by the following two-factor linear model:

$$r^{(n)}_t = \alpha^{(n)} + \beta^{(n)} mkt_t + \gamma^{(n)} ind_t + u^{(n)}_t,$$

where $a^{(n)}$ is the only difference from the single-factor model of Eq. (6) and the sample standard deviation following covariance matrix:

$$\Sigma$$

whose only difference from the single-factor model of Eq. (6) is the term involving the return $r^{(ind)}_t$ of the industry that the $n$-th stock belongs to. The stock universe is partitioned into 48 industries defined in [4], and the return for an industry is defined as the return of an equal-weighted portfolio composed of all the stocks in that industry.

The principal components model estimates the covariance matrix using several (in [5], five) largest eigenvalues $\lambda_1, \ldots, \lambda_p$ of the sample covariance matrix as follows:

$$\sum_{p=1}^{p} a_p v_p v_p^T,$$

where $v_p$ is the eigenvector of the sample covariance matrix corresponding to the eigenvalue $\lambda_p$.

The shrinkage to identity estimator [6] of the covariance matrix is a convex combination of the sample covariance and the identity matrix.

The shrinkage to market estimator proposed in [5] is a convex combination of the sample covariance and the estimate obtained from the single-factor market model. For the experiments involving this estimator, we use the authors’ code from www.ledoit.net/ole2_abstract.htm.

3. STANDARD ERRORS FOR PORTFOLIO RISK ASSUMING RETURNS ARE IID GAUSSIAN

3.1. Using portfolio risk estimates from [5]

We first assume that the monthly returns $R_1, \ldots, R_T$ of a portfolio during the time period $[1, T]$ are iid Gaussian random variables with standard deviation $\sigma$. We adopt the following commonly used definitions for the sample mean $M$ and the sample standard deviation $S$:

$$M = \frac{1}{T} \sum_{i=1}^{T} R_i$$

$$S = \sqrt{\frac{1}{T-1} \sum_{i=1}^{T} (R_i - M)^2}$$

$\sigma S$ has the chi distribution with $T - 1$ degrees of freedom [1]. Hence, the standard deviation of $S$ is [1]:

$$\sigma_S = \sigma \sqrt{1 - \frac{2}{T-1} \cdot \frac{\Gamma^2 \left(\frac{T}{2}\right)}{\Gamma^2 \left(\frac{T-1}{2}\right)}}$$

Replacing $\sigma$ with its estimate $\hat{\sigma}$, we obtain an estimate of $\sigma_S$:

$$\hat{\sigma}_S = \hat{\sigma} \sqrt{1 - \frac{2}{T-1} \cdot \frac{\Gamma^2 \left(\frac{T}{2}\right)}{\Gamma^2 \left(\frac{T-1}{2}\right)}}$$

For the values of $S$ in Table 1, and for $T = 276$ monthly returns during the testing period August 1972 through July 1995, Eq. (8) yields the standard error estimates given in parentheses in Table 2. These standard errors indicate that the differences among the portfolio risks produced by principal components, shrinkage to identity, and shrinkage to market, are statistically insignificant.

We emphasize here that all the data necessary to compute the standard errors for the portfolio risks given in Table 1 using Eq. (8) are the portfolio risks themselves and the value of $T$. Thus, our calculations given in parentheses in Table 2 are easily verifiable. Note moreover that the assumption that portfolio returns are normally distributed is overly optimistic. In fact, typical returns have heavier tails than a Gaussian distribution [8]. Hence, our estimates can be viewed as lower bounds for the actual standard errors. In Section 4, we dispense with the Gaussian assumption to construct more realistic standard error estimates.

3.2. Estimating portfolio risk ourselves

We repeat the experiments of [5]. The results are reported in Table 3. The numbers in parentheses are the standard errors estimated as in the previous section by assuming that the returns for each portfolio are iid Gaussian.
The discrepancies in the portfolio risk estimates between Tables 3 and 1 cannot be explained without knowing the exact data collection procedure used in [5], but perhaps might be due to the fact that our stock universe could be somewhat different from what was used in [5].

The conclusion is similar to that of Section 3.1: the differences among the portfolio risks produced by the industry factors, principal components, shrinkage to identity, and shrinkage to market, are statistically insignificant.

### 4. STANDARD ERRORS VIA BOOTSTRAPPING

To construct more realistic estimates of the standard errors, we use bootstrapping [8]. Suppose \( r_1, \ldots, r_T \) are the observed returns of a portfolio for times \( 1, \ldots, T \). Bootstrapping proceeds by using these observations to construct a distribution, and repeatedly resamples from this distribution. The size of each resampling is \( T \). For the \( i \)-th resampling, the sample standard deviation \( \sigma_{\text{boot}}^i \) is computed. The standard error for the risk is then estimated as the sample standard deviation of the sequence \( \sigma_{\text{boot}}^1, \ldots, \sigma_{\text{boot}}^I \), where \( I \) is the number of resamplings. The returns are assumed to be iid and have discrete uniform distribution over the observations \( r_1, \ldots, r_T \). Each resampling therefore is obtained by randomly sampling \( T \) times (with replacement) from the observations \( r_1, \ldots, r_T \). This is a standard bootstrapping approach [8]. We use \( I = 50 \) resamplings in our experiments. The results are in Table 4.

As expected, the standard error estimates obtained through bootstrapping are larger than those based on the assumption that the returns are iid Gaussian. These results suggest that the differences among the portfolio risks for the industry factors, principal components, shrinkage to identity, and shrinkage to market, are even less statistically significant than suggested by the previous section’s results.

### 5. ACKNOWLEDGMENTS

This work is based on a semester project done by the first four authors as part of a financial engineering course taught by the last author at Purdue’s School of Electrical and Computer Engineering in Spring 2010. We would like to acknowledge Purdue University Libraries for subscribing to CRSP [2] and other financial databases through WRDS [10] and thus enabling us to access the data used in this research.

### 6. REFERENCES


[10] Wharton Research Data Services (WRDS) was used in preparing this paper. This service and the data available thereon constitute valuable intellectual property and trade secrets of WRDS and its third-party suppliers.