COVARIANCE ESTIMATION AND RELATED PROBLEMS IN PORTFOLIO OPTIMIZATION

Ilya Pollak

Purdue University
School of Electrical and Computer Engineering
West Lafayette, IN 47907
USA

ABSTRACT
This overview paper reviews covariance estimation problems and related issues arising in the context of portfolio optimization. Given several assets, a portfolio optimizer seeks to allocate a fixed amount of capital among these assets so as to optimize some cost function. For example, the classical Markowitz portfolio optimization framework defines portfolio risk as the variance of the portfolio return, and seeks an allocation which minimizes the risk subject to a target expected return. If the mean return vector and the return covariance matrix for the underlying assets are known, the Markowitz problem has a closed-form solution.

In practice, however, the expected returns and the covariance matrix of the returns are unknown and are therefore estimated from historical data. This introduces several problems which render the Markowitz theory impracticable in real portfolio management applications. This paper discusses these problems and reviews some of the existing literature on methods for addressing them.

Index Terms— Covariance, estimation, portfolio, market, finance, Markowitz

1. INTRODUCTION
The return of a security between trading day \( t_1 \) and trading day \( t_2 \) is defined as the change in the closing price over this time period, divided by the closing price on day \( t_1 \). For example, the daily (i.e., one-day) return on trading day \( t \) is defined as \( (p(t) - p(t-1))/p(t-1) \) where \( p(t) \) is the closing price on day \( t \) and \( p(t-1) \) is the closing price on the previous trading day. Note that if \( t \) is a Monday or the day after a holiday, the previous trading day will not be the same as the previous calendar day.

Suppose an investment is made into \( N \) assets whose return vector is \( R \), modeled as a random vector with expected return \( \mu = E[R] \) and covariance matrix \( \Lambda = E[(R - \mu)(R - \mu)^T] \). In other words, \( R = (R^{(1)}, \ldots, R^{(N)})^T \) where \( R^{(n)} \) is the return of the \( n \)-th asset. It is assumed throughout the paper that the covariance matrix \( \Lambda \) is invertible. This assumption is realistic, since it is quite unusual in practice to have a set of assets whose linear combination has returns exactly equal to zero. Even if an investment universe contained such a set, the number of assets in the universe could be reduced to eliminate the linear dependence and make the covariance matrix invertible.

Out of these \( N \) assets, a portfolio is formed with allocation weights \( w = (w^{(1)}, \ldots, w^{(N)})^T \). The \( n \)-th weight is defined as the amount invested into the \( n \)-th asset, as a fraction of the overall investment into the portfolio: if the overall investment into the portfolio is \( $D \), and \( $D^{(n)} \) is invested into the \( n \)-th asset, then \( w^{(n)} = D^{(n)}/D \).

Therefore, by definition, the weights sum to one:
\[ w^T 1 = 1, \]
where \( 1 \) is an \( N \)-vector of ones. Note that some of the weights may be negative, signifying short positions obtained by selling borrowed units of an asset. It is assumed throughout the paper that short selling can be done freely for all \( N \) assets.

It is easily shown then that the total portfolio return is \( w^T R \). The expected portfolio return is therefore \( w^T \mu \), and the variance of the return is \( w^T \Lambda w \).

1.1. Classical Markowitz Portfolio Optimization
The classical Markowitz portfolio framework [37] defines portfolio risk as the variance of the portfolio return, and seeks a portfolio weight vector \( w \) which minimizes the portfolio risk subject to a target expected return \( \mu_{tgt} \):

Find \( w^* \) to minimize \( w^T \Lambda w \)
subject to \( w^T \mu = \mu_{tgt} \)

Using Lagrange multipliers to perform minimization (2) subject to the constraints (1) and (3) yields [37, 45]:
\[ w^* = \Lambda^{-1} m A^{-1} c, \]
where \( m = (\mu^T, 1), A = m^T \Lambda^{-1} m, \) and \( c = (\mu_{tgt}^T, 1)^T \).
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:
\[ w_{gmvp} = \frac{\Lambda^{-1} 1}{1^T \Lambda^{-1} 1}. \]

1.2. Practical Difficulties with the Framework
In practice, the expected returns and the covariance matrix of the returns are not known and are therefore estimated from historical data [14, 25, 26, 27, 32, 42, 7, 33, 2, 34, 35, 10, 53]. This introduces three well-known problems:
- Over long time periods, financial data is typically nonstationary. This limits the amount of data that can be used to meaningfully estimate the mean and the covariance of the asset return vector. On the other hand, sample covariance has many parameters and requires large amounts of data to estimate. For example, if a portfolio includes 1000 stocks, then sample covariance has roughly 500,000 parameters to be estimated. Therefore, alternative estimators of the covariance are often required.
• The optimal portfolio weights are very sensitive to the estimated means and covariances [25, 26, 27, 39, 3, 28, 9, 4]. In other words, a small change of the estimates may lead to a drastic change of portfolio weights computed by replacing the mean vector and the covariance matrix in Eqs. (4) or (5) with their estimates.

• The optimal portfolio tends to amplify large estimation errors in certain directions [25, 39]. This stems from the fact that if the variance of an asset (or a sub-portfolio of assets) is significantly underestimated and thus appears to be small, the optimal portfolio will assign a large weight to it. Similarly, a large weight will be assigned if the mean return of an asset or a sub-portfolio appears to be large as a result of being significantly overestimated. As a result, the risk of the estimated optimal portfolio is typically underpredicted and its return is overpredicted [30, 31].

In Section 2, we review a number of covariance estimation methods that have been used in the literature to overcome the first of these three problems. However, regardless of which covariance estimation method is used, the second and third problem still remain. Several methods for addressing them are reviewed in Section 3.

2. COVARIANCE ESTIMATION METHODS FOR MARKOWITZ PORTFOLIOS

Apart from using the sample covariance, the current portfolio optimization literature has three main types of approaches for estimating the covariance matrix: imposing a parametric model; bootstrapping; and shrinkage methods. To empirically evaluate a covariance estimation method, typically the following testing procedure is adopted:

1: Two parameters are selected, the length of the training window \(L\) and the length of the holding window \(K\).
2: Mean returns and the covariance matrix 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.

Using this procedure, several covariance estimators described below are evaluated in [41].

2.1. Simple Estimators

Several basic covariance estimators are often used as benchmarks in the literature.

Diagonal estimators assume that the asset returns are pairwise uncorrelated. The variances of individual asset returns are usually estimated as sample variances. An even simpler multiple-of-identity model is one of the approaches used in [34] where the estimate of the covariance matrix is a scalar multiple of the identity matrix, with the mean of the sample variances used as the diagonal entry.

The constant correlation estimator assumes that every pair of stocks in the portfolio has the same correlation coefficient [14]. For a portfolio with \(N\) assets, this model has \(N+1\) parameters: \(N\) return variances and one correlation, all of which are estimated using the corresponding sample quantities.

Both the constrained and unconstrained optimal weights require inverse covariance matrices, as evident from Eqs. (4,5). In many practical situations the number of stocks is larger than the number of historical returns per stock, resulting in a singular sample covariance matrix. The pseudoinverse method estimates the inverse covariance matrix as the pseudoinverse of the sample covariance [34].

2.2. Covariance Estimation via Linear Factor Models

Linear factor models reduce the number of model parameters by representing the return of each asset in the portfolio as a linear combination of a small number of factors such as the return of a market index, the return of an industry index, etc. Specifically, consider a portfolio with \(N\) assets whose return vector is \(\mathbf{R} = (R_1^{(t)}, \ldots, R_N^{(t)})^T\). Let \(J\) be the total number of factors used in the model, and denote their return vector by \(\mathbf{R}_f = (R_f^{(1)}, \ldots, R_f^{(J)})^T\). The corresponding linear factor model is:

\[
R_n^{(t)} = \alpha_n + \sum_{j=1}^{N} \beta_{n,j} R_f^{(j)} + U_n, \quad \text{for} \quad n = 1, \ldots, N.
\]  

Here, the random variables \(U_n\) are assumed to be zero-mean and uncorrelated both with each other and with the factor returns \(R_f\). The parameters of this model are the non-random coefficients \(\alpha_n\) and \(\beta_{n,j}\), as well as the standard deviations \(\sigma_n\) of the random variables \(U_n\).

Let \(\mathbf{\alpha} = (\alpha_1, \ldots, \alpha_N)^T\), \(\mathbf{U} = (U_1, \ldots, U_N)^T\), and let \(\mathbf{B}\) be the \(N \times J\) matrix whose \((n,j)\)-th entry is \(\beta_{n,j}\). With this notation, the \(N\) scalar equations of (6) can be written as the following vector equation:

\[
\mathbf{R} = \mathbf{\alpha} + \mathbf{B} \mathbf{R}_f + \mathbf{U}.
\]  

Since \(\mathbf{\alpha}\) and \(\mathbf{B}\) are deterministic, and since \(\mathbf{U}\) is assumed to be uncorrelated with \(\mathbf{R}_f\), taking covariance matrices of both sides of the above equation yields the following:

\[
\text{cov}(\mathbf{R}) = \mathbf{B} \text{cov}(\mathbf{R}_f) \mathbf{B}^T + \text{cov}(\mathbf{U})
\]  

Estimating the covariance matrix of \(\mathbf{R}\) therefore involves estimating the \(NJ\) entries of the matrix \(\mathbf{B}\), the \(J(J+1)/2\) distinct entries of the factor covariance matrix, and the \(N\) parameters of the diagonal covariance of \(\mathbf{U}\), for a total of \((N + J/2)(J + 1)\) parameters. If \(J << N\), this number of parameters is significantly smaller than the \(N(N + 1)/2\) parameters of the sample covariance.

For example, consider a portfolio with \(N = 1000\) assets. The sample covariance matrix for such a portfolio has \(500,500\) distinct parameters to be estimated. A linear factor model with \(10\) factors has only \((1000 + 5)(10 + 1) = 11055\) parameters. Consider the fact that one year’s worth of daily returns for 1000 assets amounts to about 250,000 data points, since there are approximately 250 trading days in a year. This is far too little data to be able to estimate the \(500,500\) distinct entries of the sample covariance, yet may be sufficient to construct reasonable estimates for the 11,055 parameters of the factor model.

2.3. Examples of Linear Factor Models

In order for a factor model to work, it must effectively capture the behavior of the returns of many assets in a portfolio using a small number of factor returns. Many research efforts have therefore focused on developing effective factor models [24]. The earliest of these [48] is a model consisting of a single market factor. A broad market index, such as S&P 500, is usually taken as a proxy for the market return.
Factors defined by industries or industry sectors are also widely used. For example, [34] uses a model with 49 factors: the market factor and 48 industry factors corresponding to the 48 industries defined in [16]. The return for an industry factor is defined as the return of an equal-weighted portfolio composed of all the stocks in that industry. This model is effectively a two-factor model, since the industry factor coefficient $\beta_{nj}$ is zero unless stock $n$ belongs to industry $j$. Hence, the sum in Eq. (6) only has two nonzero terms: one for the market factor and one for the industry that the $n$-th stock belongs to. A similar model is used in [1]: all the stocks are grouped into 15 sectors, each sector is associated to an exchange-traded fund, and the return for a sector factor is taken to be the return for of the corresponding exchange-traded fund.

In addition to clustering by industry or sector, many other ways of grouping stocks have been used to construct factors. A widely used method is to classify companies according to their size and book-to-market ratio. For example, in addition to the market factor, Fama and French [15] use two more portfolio factors as factors. One portfolio is long high book-to-market stocks (so-called value stocks) and short low book-to-market stocks (so-called growth stocks), and the other portfolio is long small companies and short large companies, as measured by the companies’ market capitalization.

The Chen-Roll-Ross model [8], in addition to the market factors, contains a number of macroeconomic factors as well as indicators from the bond market: the inflation rate, industrial production, consumption growth rate, oil prices, one-month US Treasury Bill rate (TB), return on long-term US government bonds less TB, and others.

Another widely used method for constructing factor models is to perform the Karhunen-Loève transform, also known in the literature as the principal component analysis (PCA), and only keep a small number of components that correspond to the largest eigenvalues of the covariance matrix of the return vector [43, 34, 1, 50]. Simply zeroing out the smaller eigenvalues, however, would result in a singular covariance matrix, and therefore care must be taken when applying this approach to portfolio optimization and other applications that require inverting the estimated covariance matrix [1, 50].

### 2.4. Shrinkage Methods

Shrinkage methods [34, 46] strive to achieve a compromise between the instability of the sample covariance estimator and the biases introduced by model-based estimators. A shrinkage estimator is a convex combination of the sample covariance and a so-called shrinkage target which can be, for example, any one of the estimators discussed above. The mixing weight, also called shrinkage intensity, can be obtained through cross-validation.

### 3. MITIGATING PORTFOLIO SENSITIVITY TO COVARIANCE ESTIMATES

All covariance estimation methods produce estimation errors, making their direct use in the Markowitz framework problematic. This is due to the sensitivity of the optimal portfolio weights to estimation errors, discussed above in Section 1.2. To reduce this sensitivity, two types of strategies have been used: modifying the optimization criterion and combining several portfolio weight estimates.

Examples of the former strategy are robust portfolios and norm-constrained portfolios; examples of combining weight estimates are resampled portfolios and portfolio mixtures.

1. **Robust portfolios** replace the original optimization problem (1,2,3) with various robustified versions [19, 52, 6, 18]. For example, in [19] it is proposed to select the portfolio weight vector $\mathbf{w}$ to minimize
\[
\max_{\mathbf{w} \in \mathcal{S}} \mathbf{w}^T \mathbf{A} \mathbf{w},
\]
subject to the weight normalization constraint (1) and the following constraint:
\[
\min_{\mathbf{w} \in \mathcal{M}} \mathbf{w}^T \boldsymbol{\mu} \geq \mu_{tgt},
\]
where $\mathcal{S}$ and $\mathcal{M}$ are uncertainty sets for the covariance matrix and the mean return, respectively. This framework introduces the additional difficulty of having to estimate the uncertainty sets. This is done in [19] using the confidence intervals around the estimates of the covariance and the mean.

2. **Norm-constrained portfolios** add a constraint on the norm of the portfolio weights to the optimization problem (1,2,3) [23, 22, 12, 5, 21], such as $\| \mathbf{w} \| \leq \delta$ where $\delta$ is a parameter. The basic idea is that an explicit constraint on the weights would reduce the amount of possible change of the weight vector in response to the changes of the parameter estimates.

3. **Resampled portfolios** randomize the portfolio selection procedure and compute the average weights from many randomized simulations [25, 28, 17, 47, 40].

4. **Portfolio mixtures** combine the estimated optimal portfolio with either a fixed portfolio or a portfolio which depends on a small number of estimated parameters [20, 29, 10, 11, 51, 44].

### 4. CONCLUSIONS

This paper has described several covariance estimation methods utilized in the field of portfolio optimization. Despite a large body of existing literature, many open problems remain, both in the area of designing better covariance estimators, and in developing portfolio construction algorithms which are less sensitive to parameter estimation errors and are hence more practicable.

### 5. REFERENCES


---

1 Market capitalization is equal to the price per share times the number of shares outstanding.