9

Portfolio Selection: "Optimizing" an Error*

Key words: estimation risk, simulation, optimization, asset allocation

Markowitz mean-variance theory (MV) provides a classic solution to the portfolio selection problem. The risk of a portfolio (as measured by the variance of its return) can be reduced by combining assets whose returns are imperfectly correlated. Diversification, however, is not boundless.

Implementation of this approach nonetheless requires knowledge of both the expected returns on all assets comprised in a portfolio and their covariances; an information set which by definition is not available. A common way to circumvent this problem is thus to use sample estimates of such measures in the optimization procedure (the so-called plug-in approach). It follows that a prototypical investor is not only exposed to market risk, but also to estimation risk. The latter can, therefore, be defined as the loss of utility which arises from forming portfolios on the basis of sample estimates rather than true values. Clearly, even if the true moments of the asset return distributions of a portfolio were known with certainty, MV optimized portfolios would not beat other portfolios in every future investment period, since return realizations usually differ from their expected values. However, over an appropriately large investment period, MV would provide on average the optimal portfolio composition.

The aim of this case is to discuss a few relevant problems caused by the exposure to estimation risk when investors deal with portfolio selection and illustrate the resampling technique as originally proposed by Michaud (1998).[1] By relying on a statistical view of MV optimization, this method leads to a better understanding of estimation risk exposure. Sample measures provide the initial moment estimate to a

* with Giovanna Boi, Riccardo Grassi and Alessandra Palmieri.

[1] Resampled Efficiency optimization was co-invented by Richard Michaud and Robert Michaud, U.S. patent 6,003,018, worldwide patents pending. New Frontier Advisors, LLC (NFA) is exclusive worldwide licensee.

multivariate normal distribution which is used to generate asset returns. A number of independent draws are then sorted out of this multivariate population of returns in order to simulate new return series. For each resampled series, sample moments are calculated and corresponding efficient frontier portfolios are computed. The dispersion in the asset allocation of simulated portfolios comes from the estimation risk that can affect the MV frontier constructed using the plug-in approach. For any given level of expected return, an average of the portfolio weights over the simulated portfolios yields the resampled efficient portfolios. Following the procedure in Herold and Maurer (2002), the resampling approach is empirically tested and in-sample and out-of-sample performances are compared with plug-in approach results. A detailed discussion of the pros and cons of the resampling procedure can be found in Scherer (2002), whilst alternative procedures for coping with estimation risk are presented in great detail in Meucci (2005) and in Brandt (2006). Other important references are Best and Grauer (1991), Brandt (2006), Britten-Jones (1999), Chopra and Ziemba (1993) and Jobson and Korkie (1980). Extensions to dynamic portfolio strategies have been proposed by Lacoste, El Karoui and Jeanblanc (2005), Karatzas et al. (1986), Karatzas, Lehoczky and Shreve (1987), Merton (1971), Portait, Bajeux-Besnainou and Jordan (2001, 2003), and Portait and Nguyen (2002).

The structure of this case is as follows. In Sect. 9.1, classic MV portfolio optimization is presented together with major issues – such as lack of diversification across assets and instability of the optimal portfolio both in time and along the efficient frontier – which may arise from exposure to estimation risk. Section 9.2 details the resampling technique. Section 9.3 describes the MATLAB® functions we used to perform resampling. Section 9.4 presents the results of the in-sample/out-of-sample analysis.

9.1 Problem Statement

In modern finance, Markowitz's MV portfolio selection technique provides the paradigmatic solution to the problem of optimally allocating capital among risky assets. According to this approach, in each period an investor chooses a portfolio $\boldsymbol{\omega} = [w_1, w_2, \ldots, w_K]$, such that portfolio variance is minimized given a predetermined level m of expected return. Therefore the investor's problem, assuming no short-selling, may be summarized as follows:

$$\begin{aligned} &\min_{\omega} \boldsymbol{\omega}\boldsymbol{\Sigma}\boldsymbol{\omega}' \\ &\text{sub:} \quad \boldsymbol{\omega}\boldsymbol{\mu} = m, \\ &\qquad\quad \boldsymbol{\omega}\mathbf{1} = 1, \\ &\qquad\quad \boldsymbol{\omega} \geq \mathbf{0}, \end{aligned} \tag{9.1}$$

where $\boldsymbol{\mu}$ is the $(K \times 1)$ vector of expected returns, Σ the $(K \times K)$ variance–covariance matrix of returns, $\mathbf{1}$ the $(K \times 1)$ vector with all elements equal to 1. Thus, in each period the investor trades off portfolio expected return with portfolio variance. The minimum variance frontier represents the combination variance–expected returns, constructed considering portfolios that have minimum variance for a given level of expected return. The efficient frontier is the upward sloping portion of the minimum variance frontier and every investor will choose a portfolio on the efficient frontier on the basis of his personal attitude towards risk. A strongly risk averse agent will prefer low variance and low expected return efficient portfolios. A less risk averse agent will choose higher expected return and therefore riskier portfolios.

The inputs to the classical portfolio selection model are the expected return vector $\boldsymbol{\mu}$ and the variance–covariance matrix $\boldsymbol{\Sigma}$. However, since these parameters are not known with certainty, they need to be estimated from sample data, thereby exposing the asset allocation choice to an estimation risk. The optimization problem then becomes

$$\begin{aligned} &\min_{\omega} \boldsymbol{\omega}\hat{\boldsymbol{\Sigma}}\boldsymbol{\omega}' \\ &\text{sub:} \quad \boldsymbol{\omega}\hat{\boldsymbol{\mu}} = m, \\ &\qquad\quad \boldsymbol{\omega}\mathbf{1} = 1, \\ &\qquad\quad \boldsymbol{\omega} \geq \mathbf{0}, \end{aligned} \tag{9.2}$$

where $\hat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$ represent the estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Therefore, estimation risk refers to the difference between the optimal solution to (9.1) and the optimal solution to (9.2). As an example, suppose we build a portfolio using assets with equal expected returns and variances–covariances. If, because of estimation errors, sample estimates differ across assets, MV optimization will favour some assets over others, giving to the former a higher weight vis-à-vis the latter. As a result, over-weighted assets will be those with large estimated expected returns, low variances and low correlations. Alas, these assets are likely to be the ones most affected by estimation errors!

To summarize, the consequences of estimation risk on MV optimized portfolios are of three kinds:

(a) *Low degree of diversification*. MV portfolios often involve very extreme positions. In particular, as the number of assets grows, the weight on each single asset does not tend to zero as suggested by a naive notion of diversification;
(b) *Sudden shifts* in the allocation of the optimal weights along the efficient frontier, i.e., the composition of the optimal portfolio is very different for individuals that differ slightly in their attitude towards risk;
(c) *High sensitivity* of portfolio weights to small variations in expected returns. Since little changes in expected returns can completely alter the composition of MV optimal portfolios, while modifications in the variance–covariance matrix have smaller impact, it follows that errors in the sample estimates of expected return have great bearing on allocation choices. For instance, Chopra and Ziemba (1993) find that errors in mean estimates are about ten times as important as errors in variance estimates. Errors in variance estimates are in turn about twice as important as errors in covariance estimates.

As consequence, MV optimized efficient portfolios constructed using sampled means and sampled variance–covariance matrices from a given population generally score badly once their performance is verified using out of sample data. In other words, their *ex-post* reward/risk ratio is lower than expected.

9.2 Model and Solution Methodology

In order to improve MV optimization and address estimation risk, Michaud (1998) proposes a statistical procedure based on a well-known statistical procedure named bootstrapping, Maddala and Li (1996). More specifically, a *resampling* procedure aimed at the construction of a *region* of statistically equivalent efficient portfolios is introduced, and the concept of *resampled efficiency* defined. In this way, the traditional curve representing the efficient frontier becomes a region the area of which may be viewed as a measure of the uncertainty affecting the construction of the efficient frontier.

More precisely, Michaud's resampling procedure requires:

1. Collecting T historical returns on a set of K asset classes, i.e. on investments such as stocks, bonds, real estate, or cash.
2. Computing sample means $\hat{\mu}$ and the sample variance–covariance matrix $\widehat{\Sigma}$.
3. Finding optimal weights for a set of Z mean-variance efficient portfolios. Target expected returns to solve problem (9.1) are fixed in the following way. Let $\mathbb{E}r_M$ be the higher, among asset classes, expected return and let $\mathbb{E}r_{GMV}$ be the expected return on the global minimum variance portfolio.[2] Then divide the range $(\mathbb{E}r_{GMV}, \mathbb{E}r_M)$ in $Z-1$ sub-intervals

$$\{\mathbb{E}r_{GMV}, \mathbb{E}r_{GMV}+\delta, \dots, \mathbb{E}r_{GMV}+(Z-1)\times\delta=\mathbb{E}r_M\},$$

where

$$\delta=\frac{\mathbb{E}r_M-\mathbb{E}r_{GMV}}{(Z-1)}.$$

The portfolio whose expected return is $\mathbb{E}r_{GMV}$ is called portfolio with rank 1, while the one whose expected return is $\mathbb{E}r_M$ is called portfolio with rank Z. Portfolios with intermediate expected return are then named accordingly.
4. Assuming that asset returns are from a multivariate normal distribution, with mean and variance–covariance matrix equal to the sample ones and estimated in Step 2.
5. Simulating N independent draws (i.e., resampling) for each asset class from the multivariate normal distribution as defined in Step 4, with each draw consisting in T random numbers for each asset class.
6. For each simulation $i=1,\dots,N$, re-estimating a new set of optimization inputs, μ^i and Σ^i, and finding Z simulated efficient portfolios (i.e., repeating N times Step 3).

[2] This means that $\mathbb{E}r_M=\max_{i=1,\dots,n}\hat{\mu}_i$. Instead, $\mathbb{E}r_{GMV}$ is the expected return of the portfolio that solves the selection problem $\min_{\omega}\omega\widehat{\Sigma}\omega'$, under the constraints $\omega\mathbf{1}=1$ and $\omega\geq\mathbf{0}$. The constraint on the target expected return is not included.

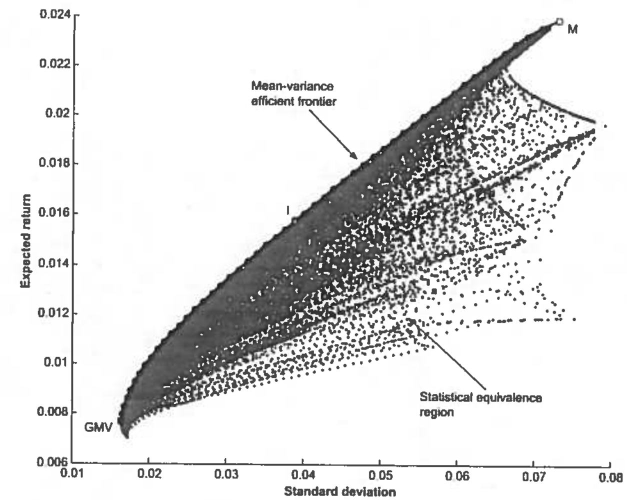

Fig. 9.1. Mean-variance efficient frontier and statistical equivalence region.

7. For each rank j, $j=1,\dots,Z$, computing the average composition across the N simulations. The Z portfolios with this average composition are called resampled portfolios of rank j.

Figure 9.1 provides a graphical representation of the statistical equivalence region and gives an immediate idea on how estimation errors can affect the determination of an efficient frontier. Note that Fig. 9.1 is constructed with the following additional steps.

(a) Calculate the expected return and the standard deviation of each of the Z mean-variance efficient portfolios using the sample mean and the sample variance–covariance matrix obtained in Step 2. Plot the MV efficient frontier.
(b) Calculate the expected return and the standard deviation of each of the $(N\times Z)$ resampled portfolios using the original sample mean and variance–covariance matrix. Plot these $(N\times Z)$ combinations variance–expected return on the same graph as the true MV efficient frontier.

Clearly, *resampled* portfolios will lie below the MV efficient frontier, as they are sub-optimal with respect to optimized portfolios based on sample estimates of the mean and the variance–covariance matrix. Nonetheless, all of these portfolios may still be considered statistically equivalent to portfolios plotted along the MV efficient frontier.

This procedure can be very useful for two main reasons. First, it provides a means of testing whether portfolios can be considered to be statistically equivalent. This may be useful in asset allocation as it may increase the stability of the optimal portfolios in time (issue (c) mentioned in the previous section) and avoid costly rebalancing. Second, the adoption of resampled portfolios may be a way of increasing portfolio diversification and stability along the frontier (problems (a) and (b) mentioned in the previous section), while maintaining coherence with the postulations of the Markowitz theory.

9.3 Implementation and Algorithm

In this section we present the MATLAB® functions used to perform the analysis and simulations. Functions used are:

- `effront.m`
- `resampfront.m`
- `simul.m`
- `stateqregion.m`
- `confregion.m`
- `resampstats.m`

Function `effront.m` computes the efficient frontier solving a standard quadratic programming problem. Optimal portfolios satisfy the expected return ranking constraint, i.e. they are equally distant in terms of expected return. A short-selling constraint is also imposed. Note that we are presenting a very simple code just to highlight basic computational steps. The reader is invited to extend the code. For instance, by adding a procedure to check if the variance–covariance matrix is semidefinite positive, or by inserting new constraints so as to limit the exposure of some asset classes.[3]

`Resampfront.m` is a loop function. It repeatedly applies `effront.m` by relying on time series generated by `mvnrnd.m`.[4] Note that the `resampfront.m` code strictly follows on a step-by-step basis the procedure described in Sect. 9.2. Using codes presented here, both our in-sample and out-of-sample analyses can be viably conducted. In-sample analysis is performed using the m-files `stateqregion.m`, `confregion.m` and `resampstats.m`. The function `stateqregion.m` jointly plots the mean-variance set and the statistical equivalence region.

[3] Note that if the global minimum variance portfolio and the maximum expected return portfolio coincide, the efficient frontier reduces to a single point. This event is unlikely but definitely possible. As the resampling procedure is based on averaging the simulated portfolios, we require that the algorithm always generates the same number of portfolios. Therefore, when only one efficient portfolio exists `effront.m` builds a $(K \times Z)$ matrix of weights which replicates the existing portfolio weights Z times. A standard mean-variance algorithm would output a single weight vector $(K \times 1)$, resulting in a dimensionality error for the resampling code.

[4] `mvnrnd.m` is the multivariate random number generator built in MATLAB®.

$$\boxed{\texttt{effront.m}} \Longrightarrow \boxed{\texttt{resampfront.m}} \Longrightarrow \boxed{\texttt{simul.m}}$$

$$\Downarrow$$

$$\boxed{\begin{array}{c}\texttt{stateqregion.m}\\ \hline \texttt{confregion.m}\\ \hline \texttt{resampstats.m}\end{array}}$$

Fig. 9.2. Flow-chart.

Command `confregion.m` finally plots confidence regions for a given set of resampled portfolios. Confidence regions are computed following the procedure described in Michaud (1998), which may be summarized as follows:

1. For each portfolio on the efficient frontier, 500 statistically equivalent portfolios are generated by simulation;
2. Resampled portfolios are computed as a mean across simulated portfolios;
3. The variance of a generic simulated portfolio, ω_S, relative to the correspondent resampled portfolio, ω_R, is defined as

$$rv = (\omega_S - \omega_R)\,\Sigma\,(\omega_S - \omega_R)';$$

4. Portfolios belonging to the $\alpha\%$ confidence region are those for which $rv \le rv^*$, where rv^* is the α percentile of the distribution of relative variances.

Function `resampstats.m` generates a structure array[5] which collects sample statistics on the distribution of portfolio weights. Function `simul.m` is a loop function used to simultaneously construct the MV frontier and perform the resampling procedures (results presented in next section are based on `simul.m`). The following flow-chart (Fig. 9.2) presents the logic relationships between functions described in this section.

To perform in-sample analysis:

1. Load a return time series, `RetSeries`, into the workspace. `RetSeries` needs to be a $(T \times K)$ matrix, where T indicates the number of returns and K the number of assets. Returns in `RetSeries` start from more recent observations going back into the past. Fix the number of portfolios `NumPortf` to be generated along the efficient frontier and the number `N` of simulations to be run.
2. Call function `resampfront.m` to generate the mean-variance and the resampled frontiers.

```
>> [Wrsp,ERrsp,SDrsp,Wmv,ERmv,SDmv,Wmv_S]
   = resampfront(RetSeries,NumPortf,N)
```

3. Call function `stateqregion.m` to generate a plot of the statistical equivalence region.

```
>> stateqregion(RetSeries, ERmv, SDmv, Wmv_S)
```

[5] Structure arrays are particular arrays having fields. Each field can store data of different type and dimension.

4. Call function `confregion.m`. This function generates a plot of the confidence region for a set of resampled portfolios – specify the portfolio set by mean of the input vector (`PortfSet`), specify the confidence level (`ConfLevel`) .

```
>> confregion(Wrsp,ERrsp,SDrsp,ERmv,SDmv,Wmv_S,
                RetSeries,PortfSet,ConfLevel)
```

5. Call function `resampstats.m` to obtain sample statistics about the distribution of portfolio weights.

```
>> [Stats] = resampstats(Wmv, Wmv_S, Wrsp, PortfSet,
                         ConfLevel)
```

Note that Steps 3–5 can be performed independently. For instance, one can decide to obtain simply confidence regions performing Steps 1, 2 and 4.

To perform out-of-sample simulation:

1. Load a return time series, `RetSeriesTotal` – postscript `Total` has been chosen to indicate a time series which spans the entire simulation period.
2. Call function `simul.m` – this function will repeatedly call mean variance and resampling procedure comparing the performance of the two methods and storing results period by period. Input data T indicates the time length of the in-sample period; N and *NumPortf* respectively indicate the number of simulations to perform in resampling procedure and the number of portfolios defining mean variance and resampled set.

```
>> [ASRmv,ASRrsp,ATOmv,ATOrsp]
   = simul(RetSeriesTotal,T,N,NumPortf)
```

9.4 Results and Comments

In this section, we provide a description of data used to perform simulations and discuss our results. In Sect. 9.4.1, we analyze the relationship between mean-variance and resampled portfolios. Section 9.4.2 presents results of the out-of-sample simulation.

Table 9.1 shows the input asset classes used together with benchmarks chosen to represent them. We consider the MSCI equity indices of six countries (Canada, US, Germany, Japan, France, UK),[6] plus Merril Lynch bond indices for the Euro and US area. All returns are expressed in US dollars. Stock Benchmark indices are downloaded from Datastream, and bond benchmarks from Bloomberg.

For each asset class we compute monthly logarithm returns for the period 1/1/85 to 8/31/02, for a total of $T = 200$ observations. We start our analysis by generating mean-variance and resampled efficient frontiers (see Figs. 9.1 and 9.3) with respect to the first 50 observations, i.e., $T = 50$. We then perform a rolling out-of-sample simulation by comparing the performance of mean-variance and resampled efficient portfolios. The in-sample period is respectively set equal to 30, 60 and 90 periods. The out-of-sample period is kept fixed and equal to $T = 100$. Further details of these simulations will be given in Sect. 9.4.2 where we describe the results obtained.

[6] MSCI provides global equity indices, which, over the last 30 years, have been the most widely used international equity benchmark indices by institutional investors. Additional information on the construction of these indices can be found at www.msci.com.

Table 9.1. Asset classes for empirical analysis

Asset class	Benchmark
Canada	MSCI Canada
US	MSCI US
Germany	MSCI Germany
Japan	MSCI Japan
France	MSCI France
UK	MSCI UK
Euro Bonds	ML Euro aggregate
US Bonds	ML US aggregate

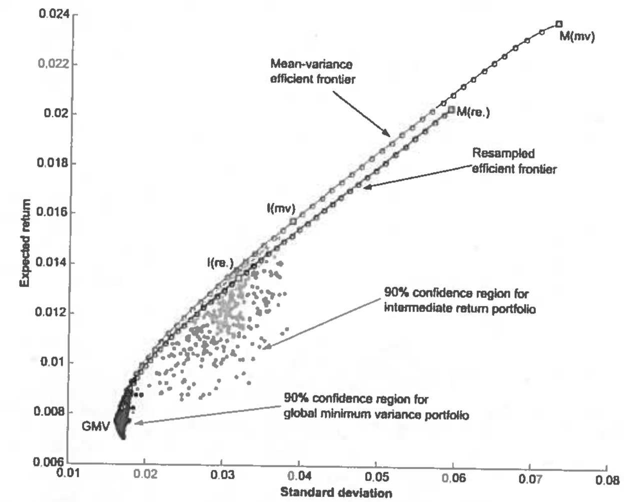

Fig. 9.3. 90% confidence regions for global minimum variance and intermediate return portfolios.