JOURNAL OF INVESTMENT MANAGEMENT, Vol. 17, No. 1, (2019), pp. 77–91

www.joim.com

PORTFOLIO OPTIMIZATION WITH NOISY COVARIANCE MATRICES

Jose Menchero[a] *and Lei Ji*[b]

In this paper, we explore the effect of sampling error in the asset covariance matrix when constructing portfolios using mean–variance optimization. We show that as the covariance matrix becomes increasingly ill-conditioned (i.e., "noisy"), optimized portfolios exhibit certain undesirable characteristics such as under-prediction of risk, increased out-of-sample volatility, inefficient risk allocation, and increased leverage and turnover. We explain these results by utilizing the concept of alpha portfolios (which explain expected returns) and hedge portfolios (which serve to reduce risk). We show that noise in the covariance matrix leads to systematic biases in the volatility and correlation forecasts of these portfolios, which in turn explains the adverse effects cited above.

1 Introduction

Mean–variance optimization, pioneered by Markowitz (1952), provides a prescription for improving the risk-adjusted performance of portfolios. The Markowitz framework identifies the *optimal* portfolio, defined as having the highest expected return per unit of risk, subject to a set of investment constraints. Three basic inputs are required for mean–variance optimization: (a) the expected returns of the assets, (b) the estimated asset covariance matrix, and (c) the set of investment constraints.

[a]Head of portfolio analytics research at Bloomberg. E-mail: jmenchero@bloomberg.net
[b]Senior quantitative researcher in portfolio analytics at Bloomberg. E-mail: lji19@bloomberg.net

The use of mean–variance optimization as a portfolio-construction tool has greatly expanded over time. Increases in computational power, coupled with advances in optimization algorithms, have now made portfolio optimization a practical tool for managing real-world portfolios. Nevertheless, given the obvious appeal of improving risk-adjusted performance, it is natural to ask why more practitioners have not yet adopted mean–variance optimization as a portfolio-construction technique.

Part of the explanation may be that some portfolio managers are reluctant to introduce advanced quantitative techniques into their investment process. However, even some quantitative practitioners remain wary of the technique due to the perception that optimized portfolios "behave badly" out-of-sample. For example, Michaud (1989) argues that mean–variance optimization often leads to non-intuitive portfolios with highly unstable weights. He further characterizes optimizers as being in a fundamental sense "error maximizers." In many cases, he claims that simple equal weighting yields superior results to mean–variance optimization.

This view is by no means universally shared. For example, Kritzman (2006) argues that while the portfolio weights may indeed be sensitive to small changes in the inputs, the risk–return profile of the portfolio—i.e., what really matters to investors—is quite robust to small changes in inputs. Moreover, Kritzman *et al.* (2010) argue that the ostensible superiority of equal-weighted portfolios arises not from any fundamental deficiency in mean–variance optimization, but rather due to estimating expected returns using short windows.

Nonetheless, even if we accept Kritzman's arguments, it is clear that legitimate concerns remain surrounding the efficacy of mean–variance optimization. The origin of these concerns is that classical mean–variance optimization implicitly assumes that the expected returns and the asset covariance matrix are known with certainty. In reality, these quantities must be *estimated* from a limited set of data, which necessarily introduces various sources of estimation error.

One such source may be termed as "modeling" error, which occurs whenever modeling assumptions do not adequately reflect reality. For instance, many models implicitly assume that the return-generating process is stationarity, whereas in reality the return dynamics may change over time. Another example of modeling error occurs in factor models when an important factor has been omitted, which may lead to underestimation of risk and sub-optimal portfolio selection, as discussed by Ceria *et al.* (2012).

Another source of estimation error is *sampling* error, which arises whenever we use a finite set of data to estimate relationships. For instance, consider a stationary time series with a true volatility σ. If we use a finite sample of T observations to estimate the volatility, the sampling error in the volatility estimate is approximately $\sigma/\sqrt{2T}$, assuming normality. Hence, sampling error increases as we use fewer observations.

In practice, there is often an optimal tradeoff between modeling error and sampling error. For instance, to mitigate the effect of non-stationary, we should assign as much weight as possible to recent observations. However, this increases sampling error. The best volatility forecasts are obtained by optimal balancing both sources of estimation error.

Previous research on estimation error in portfolio optimization can be segmented into two basic streams. The first stream focuses on estimation error in expected return estimates. For instance, Jorion (1985) highlights the pitfalls of using the sample mean for estimating expected returns. He argues that "shrinking" the expected return estimates toward some global mean provides better out-of-sample performance.

Another example of research relating to errors in expected returns is provided by Black and Litterman (1991). They first use equilibrium theory to arrive at expected returns under the assumption that the market portfolio is efficient. They then express their views in terms of deviations from the market consensus, finding that this provided more robust expected-return estimates.

The second stream of research focuses on estimation error in the asset covariance matrix. For instance, Ledoit and Wolf (2003) show that shrinking the sample asset covariance matrix toward an appropriate target[1] produces optimized portfolios with improved out-of-sample performance. The technique introduced by Ledoit and Wolf presents a practical method for mitigating the adverse effects of sampling error.

In this paper, we study the adverse effects on portfolio optimization arising from sampling error in the asset covariance matrix. In particular, we show that sampling error may give rise to several detrimental outcomes, including (a) underestimation of risk of optimized portfolios, (b) increased out-of-sample portfolio volatility, (c) inefficient allocation of the risk budget, and (d) increased portfolio leverage and turnover.

To investigate the underlying causes of these effects, we employ a conceptual framework for analyzing optimized portfolios. More specifically, we decompose the optimal portfolio into two distinct portfolios. The first is known as the *alpha* portfolio, which fully accounts for the expected return; and the second is known as the *hedge* portfolio, which has zero expected return but serves to hedge the risk of the alpha portfolio. To our knowledge, this framework has not been previously considered in the literature.

We derive several important *ex ante* properties of the alpha and hedge portfolios, and provide a simple geometric interpretation of optimal portfolios. Using this framework, we show that biases in the volatility and correlation forecasts of the alpha and hedge portfolios provide the basic mechanism to explain the out-of-sample behavior of optimized portfolios. We then present an empirical study confirming the existence of these biases. We conclude with a brief discussion on how these concepts apply to real-world investment problems.

2 Adverse effects of sampling error

In this section, we illustrate some of the adverse effects of sampling error in portfolio optimization. To study these effects, we apply mean–variance optimization to construct the minimum-volatility fully invested portfolio for a universe of the 100 largest US stocks (as of 31-Mar-2016) with complete daily return history going back to 13-Jan-1999. At the start of each day t, we estimate the asset covariance matrix $\boldsymbol{\Omega}_t^\tau$ using exponentially weighted moving averages (EWMA) with half-life parameter τ. For simplicity, we assume that the mean stock returns are negligible compared to their volatilities, which typically is indeed the case for daily stock returns. Within this approximation, the individual elements of $\boldsymbol{\Omega}_t^\tau$ are given by

$$\Omega_t^\tau(i, j) = \sum_{m=1}^{t-1} v_m^\tau r_m(i) r_m(j), \qquad (1)$$

where $r_m(i)$ is the return of stock i for day m and v_m^τ is the EWMA weight. We use approximately two years of daily data to estimate the first asset covariance matrix, which leads to start date of 27-Dec-2000 for the out-of-sample period.

The half-life parameter τ provides a convenient "knob" to adjust the degree of sampling error (or noise) in the asset covariance matrix. By reducing the half-life, we effectively use fewer observations to estimate the relationship between assets, resulting in a noisier covariance matrix.

Loosely speaking, we use the term "well-conditioned" to describe a covariance matrix in which the noise level is relatively low. Conversely, a covariance matrix with high sampling error is termed "ill-conditioned." These terms represent useful concepts, although in practice there is no clear dividing line that separates the two.

As shown by Grinold and Kahn (2000), the minimum-volatility fully invested portfolio is given by

$$\mathbf{w}_t^\tau = \frac{(\boldsymbol{\Omega}_t^\tau)^{-1}\mathbf{1}}{\mathbf{1}'(\boldsymbol{\Omega}_t^\tau)^{-1}\mathbf{1}}. \tag{2}$$

where $\mathbf{1}$ is an $N \times 1$ vector of ones and N is the total number of assets.[2] The predicted volatility of portfolio $\mathbf{w}_t^\tau$ at the start of period t is given by

$$\hat{\sigma}_t^\tau = \sqrt{\mathbf{w}_t'^\tau \boldsymbol{\Omega}_t^\tau \mathbf{w}_t^\tau}. \tag{3}$$

The return of the portfolio over period t is

$$R_t^\tau = \mathbf{r}_t' \mathbf{w}_t^\tau, \tag{4}$$

where $\mathbf{r}_t$ is the $N \times 1$ vector of realized returns for period t.

We first study how sampling error affects the accuracy of risk forecasts for optimized portfolios. A useful measure for such purposes is the *bias statistic*, which is given by the standard deviation of standardized portfolio returns. Again assuming that the mean daily returns are negligible relative to their volatility, the bias statistic is computed as

$$B_\tau = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\frac{R_t^\tau}{\hat{\sigma}_t^\tau}\right)^2}, \tag{5}$$

where T is the total number of periods. The bias statistic essentially represents the ratio of realized risk to forecast risk. Note that if the forecast volatility is equal to the *true* volatility, then the expected value of the squared bias statistic is exactly 1 by construction.[3] A bias statistic significantly greater than 1, therefore, indicates under-prediction of risk.

In Figure 1, we plot the bias statistics (left axis) of the minimum-volatility fully invested portfolios versus the half-life parameter τ. Again, these portfolios were constructed for the 100-stock universe using Equation (2). The out-of-sample period is from 27-Dec-2000 to 31-Mar-2016. For a half-life of 150 days, the bias statistic is approximately 1.3, indicating a 30 percent under-prediction of risk. While a 30 percent bias may be considered significant for certain purposes, at least the forecast is in the "right ballpark." The same cannot be said for short half-life parameters. For instance, a 10-day half-life results in a bias statistic of approximately 35, indicating a realized volatility roughly *35 times larger* than the predicted volatility. Clearly, such extreme under-forecasting would be unacceptable from a risk-management perspective.

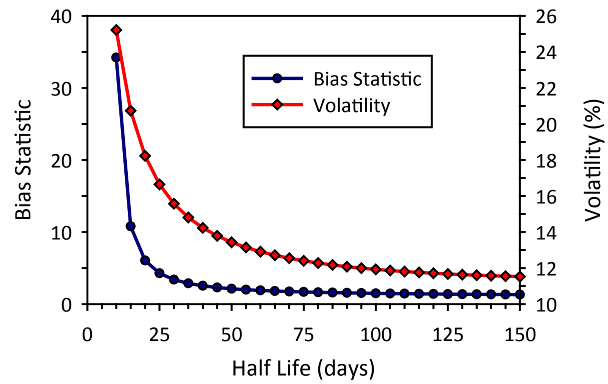

Figure 1 Bias statistics (left axis) and out-of-sample volatility (right axis) for minimum-volatility fully invested portfolios, plotted versus half-life. Results were obtained for a 100-stock portfolio with an out-of-sample testing period from 27-Dec-2000 to 31-Mar-2016, as described in the text.

Next, we consider how sampling error affects the out-of-sample volatilities of optimized portfolios. Recall that the current investment objective is to minimize portfolio volatility, subject to the full-investment constraint. The out-of-sample volatility is given by

$$\tilde{\sigma}_\tau = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(R_t^\tau)^2}, \tag{6}$$

where we have again assumed that the mean returns are negligible compared with their volatilities.

In Figure 1, we plot the out-of-sample volatilities (right axis) of optimized portfolios as a function of τ. We see that realized volatility increases significantly as the half-life parameter is decreased. For instance, using $\tau = 150$ days results in a realized volatility under 12 percent annualized versus a realized volatility of more than 25 percent for a half-life of 10 days.

To judge the merit of mean–variance optimization, we must compare the realized volatilities in Figure 1 with the volatility that would have resulted *without* mean–variance optimization. In the absence of any specific information on individual stocks, the most direct way to construct a fully invested portfolio is to simply assign equal weight to each stock, sometimes referred to as the $1/N$ portfolio. Over the sample period, the equal-weighted portfolio had a realized volatility of 16.5 percent. Hence, from Figure 1, we see that when the half-life is greater than 25 days, mean–variance optimization leads to a meaningful reduction in portfolio volatility. This result shows that mean–variance optimization can indeed improve risk-adjusted performance, as long as the covariance matrix is well-conditioned.

By contrast, when the half-life is below 25 days, excessive noise in the asset covariance matrix causes the portfolio volatility to *exceed* 16.5 percent. In other words, using an ill-conditioned covariance matrix for portfolio optimization can actually do more harm than good.

Other portfolio characteristics of interest to investors are the leverage and turnover of the optimized portfolios. The weight of asset n in portfolio $\mathbf{w}_t^\tau$ is given by

$$w_t^\tau(n) = \boldsymbol{\delta}_n' \mathbf{w}_t^\tau, \tag{7}$$

where $\boldsymbol{\delta}_n$ is an $N \times 1$ vector with 1 in entry n and zeros elsewhere. We define the turnover as the mean absolute value of the change in portfolio weights from one period to the next,[4]

$$TO_\tau = \frac{1}{T} \sum_{t=1}^{T} \sum_{n=1}^{N} |w_{t+1}^\tau(n) - w_t^\tau(n)|, \tag{8}$$

where T is the total number of periods. Similarly, leverage is defined as the mean sum of the absolute value of portfolio weights,

$$L_\tau = \frac{1}{T} \sum_{t=1}^{T} \sum_{n=1}^{N} |w_t^\tau(n)|. \tag{9}$$

In Figure 2, we plot the mean daily turnover (left axis) and the mean leverage (right axis) as a function of half-life. Again, the out-of-sample period is from 27-Dec-2000 to 31-Mar-2016. We see that both quantities rise dramatically as sampling error increases. For instance, using $\tau = 10$ days results in a mean daily turnover of nearly 250 percent, and a leverage of more than 10, implying a portfolio that is 550 percent long and 450 percent short. In practice, few investors would ever hold such portfolios.

This example illustrates an important lesson in portfolio construction using mean–variance optimization. Namely, to obtain a portfolio that

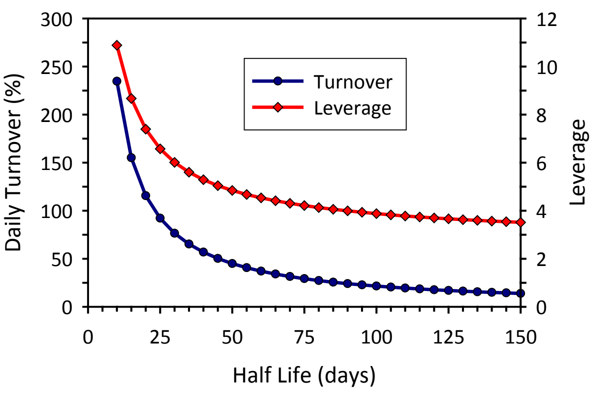

Figure 2 Daily turnover (left axis) and leverage (right axis) for minimum-volatility fully invested portfolios, plotted versus half-life. Results were obtained for a 100-stock portfolio with an out-of-sample testing period from 27-Dec-2000 to 31-Mar-2016, as described in the text.

behaves well out-of-sample, it is imperative to use a well-conditioned covariance matrix.

3 The alpha and hedge portfolios

In this section, we present a conceptual framework for representing and analyzing optimal portfolios. In particular, we decompose the optimal portfolio holdings vector into two distinct portfolios. The first portfolio, which fully explains the expected return, we term as the *alpha* portfolio. The second portfolio, which has zero expected return but serves to reduce portfolio risk, we label as the *hedge* portfolio. As we now describe, these portfolios satisfy several important properties. Details may be found in Technical Appendix A.

Let $\boldsymbol{\alpha}$ denote an $N \times 1$ vector of expected stock returns, or alphas.[5] Without loss of generality, this can be written as

$$\boldsymbol{\alpha} = a\mathbf{e}_\alpha, \tag{10}$$

where $\mathbf{e}_\alpha$ is a unit-norm vector (i.e., $\mathbf{e}'_\alpha \mathbf{e}_\alpha = 1$) and $a = \sqrt{\boldsymbol{\alpha}'\boldsymbol{\alpha}}$ is the norm of the alpha vector. From Grinold and Kahn (2000), the minimum-volatility portfolio with unit exposure to $\mathbf{e}_\alpha$ is given by

$$\mathbf{w} = \frac{\boldsymbol{\Omega}^{-1}\mathbf{e}_\alpha}{\mathbf{e}'_\alpha \boldsymbol{\Omega}^{-1}\mathbf{e}_\alpha}, \tag{11}$$

where $\boldsymbol{\Omega}$ is the $N \times N$ asset covariance matrix. The expected return of this portfolio is $\mathbf{w}'\boldsymbol{\alpha} = a$. Note that this portfolio has the minimum *ex ante* risk of *all* portfolios with the same level of expected return. Hence, portfolio $\mathbf{w}$ has the maximum expected Sharpe ratio.

Next, we write the optimal portfolio as the alpha portfolio $\mathbf{e}_\alpha$ plus the hedge portfolio $\mathbf{h}$,

$$\mathbf{w} = \mathbf{e}_\alpha + \mathbf{h}. \tag{12}$$

Since the optimal portfolio and the alpha portfolio are both known, the hedge portfolio is easily computed as $\mathbf{h} = \mathbf{w} - \mathbf{e}_\alpha$. As shown in Technical Appendix A, the hedge portfolio satisfies three key properties.

Property 1: The hedge portfolio has zero expected return. That is,

$$\boldsymbol{\alpha}'\mathbf{h} = 0. \tag{13}$$

Property 2: The hedge portfolio is uncorrelated with the overall portfolio. In other words,

$$\mathbf{w}'\boldsymbol{\Omega}\mathbf{h} = 0, \tag{14}$$

which implies that the hedge portfolio contributes zero to the risk of the optimal portfolio. This result may be interpreted as a self-consistency requirement. More specifically, for an unconstrained optimal portfolio, the risk contribution must be proportional to the expected return contribution, which is zero by Equation (13).

Property 3: The hedge portfolio is negatively correlated with the alpha portfolio. In particular, the correlation is given by

$$\rho_{\alpha h} = -\frac{\sigma_h}{\sigma_\alpha}, \tag{15}$$

where σ_h is the predicted volatility of the hedge portfolio and σ_α is the volatility of the alpha portfolio. This negative correlation provides the basic mechanism through which the hedge portfolio increases risk-adjusted performance. Namely, it reduces overall portfolio risk without altering the expected returns.

The decomposition in Equation (12) expresses the optimal portfolio as the sum of two other portfolios, which is reminiscent of the two-fund separation theorem. The similarity, however, is only superficial.

To better understand the distinction between the alpha/hedge framework and the two-fund separation theorem, it is helpful to consider the location of these portfolios with respect to the efficient frontier. In Figure 3, we plot the efficient frontier for a universe of 100 stocks under a realistic set of expected returns (i.e., alphas). This frontier represents the set of all portfolios that have

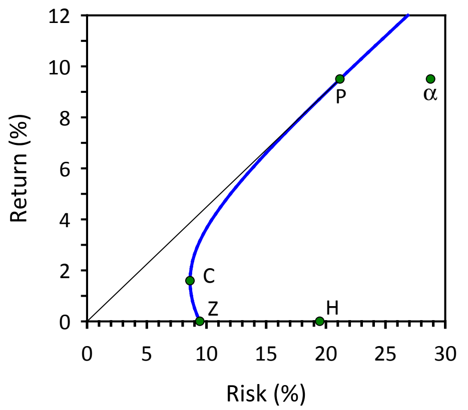

Figure 3 Efficient frontier for the fully invested portfolio. Point P is the tangent portfolio (maximum Sharpe Ratio). Point C is the minimum-volatility fully invested portfolio. Point Z is the minimum-volatility fully invested portfolio with zero beta (with respect to P). The alpha and hedge portfolios are indicated by points α and H, respectively.

minimum-volatility for a given level of expected return, subject to the full-investment constraint. The return assumptions and technical details are described in Appendix B.

Point P represents the fully invested portfolio with maximum Sharpe ratio, also known as the *tangent* portfolio. Point C represents the global minimum-volatility fully invested portfolio. Point Z represents the zero-beta portfolio, which has the lowest volatility of all fully invested portfolios that are uncorrelated to Portfolio P. The two-fund separation theorem states that all portfolios on the efficient frontier can be written as a weighted combination of any two other portfolios on the frontier. For instance, Portfolio P can be expressed as a weighted sum of Portfolios C and Z.

Equation (12) also expresses Portfolio P as the sum of two portfolios (i.e., the alpha and hedge portfolios). However, unlike the two-fund separation theorem, neither the alpha portfolio (α) nor the hedge portfolio (H) lies on the efficient frontier, as shown in Figure 3. Note that the alpha portfolio has the same expected return as the tangent portfolio, but has higher risk. The hedge portfolio has zero beta with respect to Portfolio P, but higher risk than Portfolio Z. Finally, note that in our case the alpha portfolio is 177 percent net long, while the hedge portfolio is 77 percent net short.

The alpha/hedge decomposition is also distinct from the "optimal orthogonal portfolio" concept described by MacKinlay (1995). In his framework, he defines a limited set of factor portfolios that explain part—but not all—of the expected asset returns. As a result, the tangent portfolio cannot be written as a weighted combination of these factor portfolios. However, he shows that there exists a unique portfolio (known as the optimal orthogonal portfolio) that can be combined with the original factor portfolios to arrive at the tangent portfolio. This optimal orthogonal portfolio is uncorrelated with the original factor portfolios, but not with the *tangent* portfolio. By contrast, in our formulation, the hedge portfolio is uncorrelated with the tangent portfolio and negatively correlated with the alpha portfolio. Hence, the hedge portfolio is not the same as the optimal orthogonal portfolio.

Scherer (2015) describes a framework that comes closest in spirit to the alpha/hedge decomposition. He considers an inefficient benchmark and an active manager who employs a set of alpha forecasts to outperform the benchmark. The active portfolio is constrained to be dollar neutral. Therefore, the total portfolio (i.e., benchmark plus active portfolio) is fully invested. He shows that none of these fully invested portfolios are on the efficient frontier. However, he shows that all of these portfolios can be combined with a single special add-on portfolio to arrive at all portfolios on the efficient frontier. The add-on portfolio has the following properties: (a) it is a zero-investment portfolio, (b) it has zero expected return, and (c) it is uncorrelated with all portfolios on the efficient frontier.

Our hedge portfolio H is similar to Scherer's add-on portfolio in that it also has zero expected return. Moreover, it can be combined with another portfolio (the alpha portfolio) to arrive at a point on the efficient frontier (maximum Sharpe ratio). Finally, the hedge portfolio is uncorrelated with the tangent portfolio. While there are some similarities, there are also important differences. First, while the add-on portfolio is zero investment, the hedge portfolio is not. Second, whereas the add-on portfolio is uncorrelated with *all* portfolios on the efficient frontier, the hedge portfolio is only uncorrelated with the tangent portfolio. Finally, whereas the weights of the alpha portfolio are proportional to the expected returns, the weights of the starting portfolio in Scherer's case are determined by portfolio optimization.

4 Geometric interpretation (*ex ante*)

In Figure 4, we present a geometric interpretation of mean–variance optimization in terms of alpha and hedge portfolios. The hypotenuse of the triangle labeled ABC represents the alpha portfolio, whose predicted volatility $\hat{\sigma}_\alpha$ is given by the length of line AB. Similarly, the predicted volatility $\hat{\sigma}_h$ of the hedge portfolio is given by the length of the line denoted BC. Finally, the predicted volatility $\hat{\sigma}_P$ of the optimal portfolio is given by the length of the line labeled AC. Since the hedge portfolio is uncorrelated (*ex ante*) with the optimal portfolio, lines AC and BC are perpendicular.

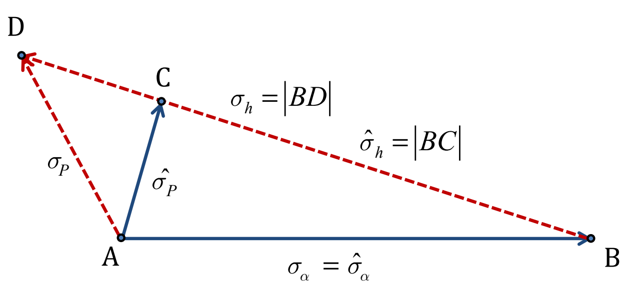

Figure 4 Risk triangle under the assumption that we correctly estimate the correlation between the hedge and alpha portfolios, but underestimate the volatility of the hedge portfolio.

Portfolio optimization is essentially equivalent to the task of finding the unique portfolio that has zero alpha and maximum negative correlation with the alpha portfolio. The correlation between these portfolios is given by the cosine of the angle between lines AB and BC, as shown in Figure 4. Indeed, since these lines are pointed in nearly opposite directions, the correlation appears strongly negative. As a result, in Figure 4, the predicted volatility of the optimal portfolio is a small fraction of the volatility of the alpha portfolio, *even though the two portfolios have the same expected return*. Consequently, the optimal portfolio has much higher risk-adjusted performance than the alpha portfolio.

5 Geometric interpretation with sampling error

Previously, we have seen that sampling error in the asset covariance matrix leads to under-prediction of risk of optimized portfolios. In this section, we present two possible explanations for this effect. One possibility is that sampling error causes the volatility forecast of the hedge portfolio to be *too low*. The second possibility is that the estimated correlation between the alpha and the hedge portfolio is *too negative*.

In Figure 4, we consider the impact of a hypothetical under-prediction of risk for the hedge portfolio. To isolate the effect of a volatility bias, we assume that the estimated correlation $\hat{\rho}_{\alpha h}$ between the alpha portfolio and the hedge portfolio is equal to the *true* correlation. The *predicted* volatility $\hat{\sigma}_h$ of the hedge portfolio is given by the length of line BC. However, since we suppose that the estimated volatility is too low, the *true* volatility σ_h of the hedge portfolio must be higher, as represented by the length of line BD.

The optimal portfolio consists of the alpha portfolio plus the hedge portfolio. Hence, the predicted volatility $\hat{\sigma}_P$ of the optimal portfolio is given by the length of the line AC, whereas the true volatility σ_P is given by the length of line AD. As shown in Figure 4, the true volatility of the optimal portfolio is significantly greater than the estimated volatility. Furthermore, whereas the *ex ante* correlation between the optimal and hedge portfolios is zero, the true correlation is positive. This implies that the hedge portfolio contributes *positively* to the risk of the optimal portfolio, even though it contributes *zero*to the expected return. In other words, it leads to an inefficient allocation of the risk budget.

The second possible explanation is that sampling error causes the estimated correlation between the alpha and hedge portfolios to be systematically too negative. In Figure 5, the estimated correlation between the alpha and hedge portfolios is given by $\cos(\hat{\theta}_h)$. Hence, according to the asset covariance matrix, the hedge portfolio is pointed in the direction of line BC. If the true correlation is less negative, as given by $\cos(\theta_h)$, then the hedge portfolio actually points in the direction of line BD.

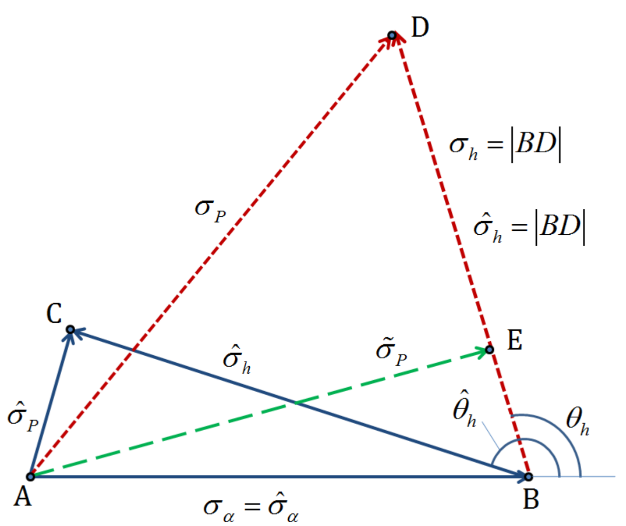

Figure 5 Risk triangle under the assumption that we correctly estimate the volatility of the hedge portfolio, but overestimate the magnitude of the correlation between the hedge and alpha portfolios.

In Figure 5, we consider the effect of such a bias in the correlation estimates. To isolate the effect of the correlation bias, we assume that the volatility of the hedge portfolio has been correctly estimated. Hence, lines BC and BD are of equal length. According to the *ex ante* estimates, the volatility of the optimal portfolio is given by the length of line AC. The true volatility, however, is much larger, given by the length of line AD. Hence, this bias in the correlation forecast would also explain the underestimation of risk of optimized portfolios. Furthermore, note that lines AD and BD now indicate positive correlation. Hence, the hedge portfolio contributes *positively* to the risk of the optimal portfolio, even though it contributes *zero* to the expected return. Again, this represents an inefficient allocation of the risk budget.

We have considered the effect of such biases in isolation. In practice, of course, both effects may occur simultaneously. It should be clear from Figure 5 that with both biases acting simultaneously, the under-forecasting of risk would only be exacerbated.

Finally, it is interesting to consider what the true optimal portfolio would be, given that the true correlation is $\cos(\theta_h)$. In this case, the optimal portfolio would place a much smaller weight on the hedge portfolio (given by the length of line BE). Therefore, the volatility of the true optimal portfolio would be given by the length of line AE. Note that line AE is perpendicular to line BE, so that the risk budget is efficiently allocated.

6 Empirical results

In the previous section, we showed that two types of biases may explain the under-prediction of risk of optimized portfolios. First, the predicted volatility of the hedge portfolio may be too low. Second, the predicted correlation between the hedge and alpha portfolios may be too negative.

In this section, we empirically test for the existence of such biases. For each stock n in our 100-stock universe, we form the minimum volatility portfolio with unit exposure to the particular stock,

$$\mathbf{w}_{nt}^{\tau} = \frac{(\boldsymbol{\Omega}_t^{\tau})^{-1}\boldsymbol{\delta}_n}{\boldsymbol{\delta}_n'(\boldsymbol{\Omega}_t^{\tau})^{-1}\boldsymbol{\delta}_n}, \tag{16}$$

where $\boldsymbol{\delta}_n$ is an $N \times 1$ vector with 1 in entry n and zeros elsewhere, and $N = 100$ is the total number of stocks. Hence, portfolio $\mathbf{w}_{nt}^{\tau}$ has a weight of 100 percent in stock n, while the weights of the other 99 stocks are designed to minimize overall portfolio risk. From Equation (12), the hedge portfolio is given by

$$\mathbf{h}_{nt}^{\tau} = \mathbf{w}_{nt}^{\tau} - \boldsymbol{\delta}_n. \tag{17}$$

The return of the hedge portfolio is given by

$$R_{nt}^{h\tau} = \mathbf{r}_t' \cdot \mathbf{h}_{nt}^{\tau}, \tag{18}$$

where $\mathbf{r}_t$ is the $N \times 1$ vector of realized returns for period t. The predicted volatility of the hedge portfolio is

$$\hat{\sigma}_{nt}^{h\tau} = \sqrt{\mathbf{h}_{nt}'^{\tau}\boldsymbol{\Omega}_t^{\tau}\mathbf{h}_{nt}^{\tau}}. \tag{19}$$

Therefore, the bias statistic of the hedge portfolio is given by

$$B_n^{h\tau} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\frac{R_{nt}^{h\tau}}{\hat{\sigma}_{nt}^{h\tau}}\right)^2}. \tag{20}$$

In Figure 6, we plot the mean bias statistics (averaged across all 100 sets of portfolios) as a function of the half-life parameter for the optimal, hedge, and alpha portfolios. First, we note that the bias statistic of the alpha portfolio is close to 1 for all ranges of half-life parameters, indicating an unbiased risk forecast. This is reasonable, given that the alpha portfolio is independent of the covariance matrix. In fact, in this case, the alpha portfolio corresponds simply to an individual stock.

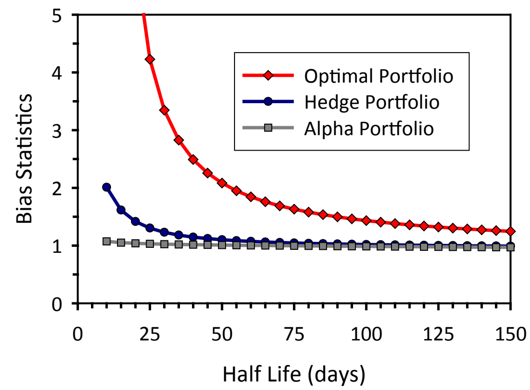

Figure 6 Bias statistics for the optimal, hedge, and alpha portfolios, plotted versus half-life parameter. Results were obtained for a 100-stock portfolio with an out-of-sample testing period from 27-Dec-2000 to 31-Mar-2016, as described in the text.

By contrast, Figure 6 shows that the covariance matrix produces biased volatility forecasts for the hedge portfolio. However, for most values of the half-life parameter, the bias is practically negligible. Nevertheless, for shorter half-life parameters, the bias can become quite significant. For instance, using a half-life of 10 days causes the risk to be under-forecast by a factor of 2.

In Figure 6, we see that the risk of the optimal portfolio is under-forecast for all ranges of half-life parameters shown here. For instance, using a half-life parameter of 50 days leads to a mean bias statistic of approximately 2. Since the bias statistic of the hedge portfolio is very close to 1 in this case, it implies that the observed under-prediction of risk is explained almost entirely by the bias in the correlation forecast.

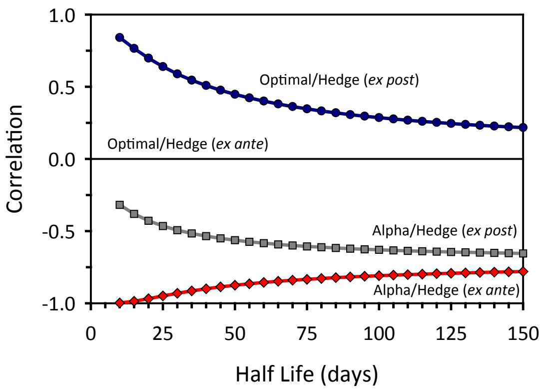

Figure 7 Portfolio correlations versus half-life parameter. Results were obtained for a 100-stock portfolio with an out-of-sample testing period from 27-Dec-2000 to 31-Mar-2016, as described in the text.

To confirm this explanation, in Figure 7 we plot *ex ante* and *ex post* correlations as a function of half-life parameter. Mean *ex post* correlations were averaged across all 100 sets of portfolios, whereas mean *ex ante* correlations were averaged across time and portfolios.

The first feature we observe in Figure 7 is that the *ex ante* correlation between the alpha and hedge portfolios is systematically biased. That is, the realized correlation is always less negative than the predicted correlation. Moreover, we see that the bias increases as the half-life becomes shorter. For example, using a half-life of 10 days produces a correlation forecast very close to -1 (implying a near-perfect hedge), whereas the realized correlation was only -0.3. In other words, as the covariance matrix becomes more ill-conditioned, the hedge appears better (*ex ante*), but is in fact increasingly poor (*ex post*).

Next, we consider the correlation between the optimal and hedge portfolio. The *ex ante* correlation between these two portfolios is exactly zero. The realized correlation, however, is always positive. Furthermore, the *ex post* correlation increases as the half-life is reduced. This means that when the covariance matrix is ill-conditioned, the hedge portfolio contributes significantly to portfolio risk, although it contributes zero to expected return.

To further illustrate these effects, in Figure 8 we present a scatterplot of the *ex post* correlation between the alpha and hedge portfolio (for each stock) versus the average *ex ante* correlation for two different half-life parameters. This plot starkly illustrates the systematic biases in the correlation forecasts. The good news is that in every single case, the realized correlations are negative. This means that the covariance matrix succeeds in identifying real hedges. The bad news is that for each of 100 portfolios, the realized correlation is less negative than the mean predicted correlation (i.e., all points lie above the 45-degree line). Also note that the 150-day half-life produces realized correlations that are more negative than for the 25-day half-life. This means that the long half-life covariance matrix produced more effective out-of-sample hedges than the 25-day model. By contrast, the *ex ante* correlations for the 25-day

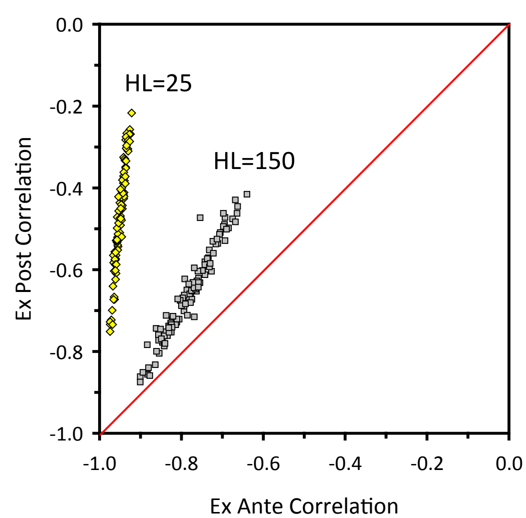

Figure 8 Scatterplot of *ex ante* and *ex post* correlations between the alpha and hedge portfolios. Results were obtained for a 100-stock portfolio with an out-of-sample testing period from 27-Dec-2000 to 31-Mar-2016, as described in the text.

half-life are more negative. This result is consistent with Figure 7. That is, a well-conditioned matrix is more modest about what it can achieve, but comes closer to reaching its objective in reality. By contrast, an ill-conditioned covariance matrix is overly optimistic about the quality of the hedge portfolio, but these hedges fail to perform effectively out-of-sample.

7 Discussion

We have demonstrated the importance of using a well-conditioned covariance matrix when using mean–variance optimization. For the small-scale example considered here (i.e., 100 stocks), this could be easily achieved by simply selecting a long half-life. For larger-scale problems, however, choosing a long half-life may not be a viable option. For example, suppose one tries to directly compute the asset covariance matrix for the constituents of the Russell 3000 index. If we use daily returns, we would need at least 3000 days (nearly 12 years) to merely avert a singular covariance matrix. In reality, we would need several times that much data (perhaps 50 years) to obtain a well-conditioned covariance matrix. This presents two serious obstacles. First, such long histories are typically not available. Second, even if the deep history were available, it would be inadvisable to use since data that is several decades old has little to do with current market conditions.

To overcome this investment obstacle, Rosenberg (1974) pioneered the use of factor models. He posited that security returns were driven by a parsimonious set of K common factors, plus an idiosyncratic component for each of the N securities. The brilliance of this approach is that the $N \times N$ asset covariance matrix could be reliably estimated using only the $K \times K$ factor covariance matrix, plus an estimate of the idiosyncratic volatility for each stock. Although the US equity market may contain thousands of stocks, several dozen factors may suffice to capture the cross-sectional variation in stock returns. In this way, a well-conditioned covariance matrix can be estimated using only a relatively small set of recent data.

Special care must be taken when estimating multi-asset risk models. These models generally span multiple markets and multiple asset classes and may have thousands of detailed factors covering very large universes. For example, the Bloomberg multi-asset risk model (MAC2) contains nearly 2,000 factors covering several million assets. To estimate the factor covariance matrix using weekly data would again require several decades of history. As before, such deep histories are typically unavailable and in any case have little to do with current market conditions.

As described by Menchero and Ji (2016), special techniques are required to ensure a well-conditioned covariance matrix for multi-asset risk models. In essence, they blended the sample factor covariance matrix with the factor covariance matrix estimated from a multi-factor model. That is, they effectively used a "factor-of-factors model" as the shrinkage target. This approach is similar to Ledoit and Wolf (2003), except for the difference in shrinkage target.

8 Summary

This paper focuses on understanding the role of sampling error in the asset covariance matrix when constructing portfolios using mean–variance optimization. By adjusting the half-life parameter, we were able to control the level of noise. We showed that ill-conditioned covariance matrices lead to several detrimental effects, such as under-forecasting of risk, increased out-of-sample volatility, increased leverage and turnover, and inefficient allocation of the risk budget.

We presented a conceptual framework for analyzing optimal portfolios by decomposing them into alpha and hedge portfolios. We showed that ill-conditioned covariance matrices lead to systematic biases in the volatility and correlation forecasts of these portfolios, which in turn explains the detrimental effects observed empirically. By contrast, these adverse effects are largely mitigated by using a well-conditioned covariance matrix, which can lead to material improvements in out-of-sample performance.

Appendix A

Let $\boldsymbol{\alpha} = a\mathbf{e}_\alpha$ be the $N \times 1$ vector of expected returns. Without loss of generality, as assume that $\mathbf{e}_\alpha'\mathbf{e}_\alpha = 1$. The minimum-volatility portfolio with unit exposure to $\mathbf{e}_\alpha$ is given by

$$\mathbf{w} = \frac{\boldsymbol{\Omega}^{-1}\mathbf{e}_\alpha}{\mathbf{e}_\alpha'\boldsymbol{\Omega}^{-1}\mathbf{e}_\alpha}, \tag{A1}$$

which we write as the sum of the alpha portfolio $\mathbf{e}_\alpha$ and the hedge portfolio $\mathbf{h}$

$$\mathbf{w} = \mathbf{e}_\alpha + \mathbf{h}. \tag{A2}$$

Note that the hedge portfolio is obtained by simply taking the difference between the optimal portfolio and the alpha portfolio, both of which are known.

The first property that we wish to prove is that the hedge portfolio has zero expected return. This is equivalent to saying that it is orthogonal to the alpha portfolio. The expected return of the hedge portfolio is given by

$$\boldsymbol{\alpha}'\mathbf{h} = a\mathbf{e}_\alpha'\left(\frac{\boldsymbol{\Omega}^{-1}\mathbf{e}_\alpha}{\mathbf{e}_\alpha'\boldsymbol{\Omega}^{-1}\mathbf{e}_\alpha} - \mathbf{e}_\alpha\right), \tag{A3}$$

which can be rewritten as

$$\boldsymbol{\alpha}'\mathbf{h} = a(1 - \mathbf{e}_\alpha'\mathbf{e}_\alpha). \tag{A4}$$

Since $\mathbf{e}_\alpha'\mathbf{e}_\alpha = 1$, it immediately follows that $\boldsymbol{\alpha}'\mathbf{h} = 0$, which proves Property 1.

Next, we wish to show that the hedge portfolio is uncorrelated with the optimal portfolio. The covariance of these two portfolios is

$$\mathbf{w}'\boldsymbol{\Omega}\mathbf{h} = \left(\frac{\mathbf{e}_\alpha'\boldsymbol{\Omega}^{-1}}{\mathbf{e}_\alpha'\boldsymbol{\Omega}^{-1}\mathbf{e}_\alpha}\right)\boldsymbol{\Omega}\left(\frac{\boldsymbol{\Omega}^{-1}\mathbf{e}_\alpha}{\mathbf{e}_\alpha'\boldsymbol{\Omega}^{-1}\mathbf{e}_\alpha} - \mathbf{e}_\alpha\right), \tag{A5}$$

which simplifies to

$$\mathbf{w}'\boldsymbol{\Omega}\mathbf{h} = \frac{1}{\mathbf{e}_\alpha'\boldsymbol{\Omega}^{-1}\mathbf{e}_\alpha}\left(\frac{\mathbf{e}_\alpha'\boldsymbol{\Omega}^{-1}\mathbf{e}_\alpha}{\mathbf{e}_\alpha'\boldsymbol{\Omega}^{-1}\mathbf{e}_\alpha} - \mathbf{e}_\alpha'\mathbf{e}_\alpha\right). \tag{A6}$$

Hence, it follows that $\mathbf{w}'\boldsymbol{\Omega}\mathbf{h} = 0$, which proves Property 2.

The third property is that the hedge portfolio is negatively correlated with the alpha portfolio. The variance of the optimal portfolio is

$$\sigma_w^2 = \sigma_\alpha^2 + \sigma_h^2 + 2\sigma_\alpha\sigma_h\rho_{\alpha h}. \tag{A7}$$

Since the hedge portfolio is uncorrelated with the optimal portfolio, the Pythagorean theorem applies

$$\sigma_\alpha^2 = \sigma_w^2 + \sigma_h^2. \tag{A8}$$

Substituting Equation (A8) into Equation (A7), and solving for $\rho_{\alpha h}$, we find

$$\rho_{\alpha h} = -\frac{\sigma_h}{\sigma_\alpha}, \tag{A9}$$

which proves the third Property.

Appendix B

Let $\mathbf{w}$ denote an $N \times 1$ vector of portfolio weights, and let $\boldsymbol{\Omega}$ denote the $N \times N$ asset covariance matrix. Our objective is to minimize the portfolio variance subject to a set of K equality constraints,

$$\min(\mathbf{w}'\boldsymbol{\Omega}\mathbf{w}) \quad \text{subject to } \mathbf{A}\mathbf{w} = \mathbf{b}, \tag{B1}$$

where $\mathbf{A}$ is the $K \times N$ constraint matrix, and $\mathbf{b}$ is a $K \times 1$ vector of constraint parameters.

For instance, suppose that we want to find the minimum variance portfolio with fixed alpha α_P, subject to the full-investment constraint. In this case, the constraint equation is given by

$$\begin{bmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_N \\ 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix} = \begin{bmatrix} \alpha_P \\ 1 \end{bmatrix}. \quad \text{(B2)}$$

To find the optimal portfolio subject to constraints, we use the method of Lagrange multipliers. We define the function

$$L(\mathbf{w}, \boldsymbol{\lambda}) = \mathbf{w}'\boldsymbol{\Omega}\mathbf{w} - (\mathbf{w}'\mathbf{A}' - \mathbf{b}')\boldsymbol{\lambda}, \quad \text{(B3)}$$

where $\boldsymbol{\lambda}$ is a $K \times 1$ vector of Lagrange multipliers. Setting the partial derivatives with respect to $\mathbf{w}$ equal to zero, we find

$$\mathbf{0}_N = 2\boldsymbol{\Omega}\mathbf{w} - \mathbf{A}'\boldsymbol{\lambda}, \quad \text{(B4)}$$

where $\mathbf{0}_N$ is an $N \times 1$ vector of zeros. Taking the derivative of Equation (B3) with respect to $\boldsymbol{\lambda}$, we obtain

$$\mathbf{0}_K = \mathbf{A}\mathbf{w} - \mathbf{b}, \quad \text{(B5)}$$

where $\mathbf{0}_K$ is an $K \times 1$ vector of zeros. We now have $K + N$ equations to solve for $K + N$ unknowns. Using Equation (B4) to solve for $\mathbf{w}$, we obtain

$$\mathbf{w} = \frac{1}{2}\boldsymbol{\Omega}^{-1}\mathbf{A}'\boldsymbol{\lambda}. \quad \text{(B6)}$$

Substituting Equation (B6) into Equation (B5), we solve for $\boldsymbol{\lambda}$

$$\boldsymbol{\lambda} = 2(\mathbf{A}\boldsymbol{\Omega}^{-1}\mathbf{A}')^{-1}\mathbf{b}. \quad \text{(B7)}$$

Finally, substituting Equation (B7) into Equation (B6), we find

$$\mathbf{w} = \boldsymbol{\Omega}^{-1}\mathbf{A}'(\mathbf{A}\boldsymbol{\Omega}^{-1}\mathbf{A}')^{-1}\mathbf{b}. \quad \text{(B8)}$$

To generate the efficient frontier shown in Figure 3, we imposed the same two constraints as in Equation (B2): i.e., fixed expected portfolio return α_P, and full-investment constraint. Construction of the efficient frontier also requires an asset covariance matrix and a set of expected returns. The asset covariance matrix $\boldsymbol{\Omega}$ was computed using EWMA with a 150-day half-life for the 100 largest US stocks as of 31-Mar-2016. To obtain the expected returns, we treated the equal-weighted portfolio as the benchmark, which we assumed to have a 5.0 percent return premium. We also assumed that the stocks had "exceptional returns" δ_n that were drawn from a normal distribution of mean zero and standard deviation of 30 bps. The expected returns (alphas) were thus given by

$$\alpha_n \equiv E[r_n] = \beta_n E[R_B] + \delta_n, \quad \text{(B9)}$$

where the betas with respect to the benchmark were computed using the asset covariance matrix $\boldsymbol{\Omega}$.

Notes

1. This is known as the "prior" in Bayesian statistics.
2. If we assume that all stocks have the same expected returns, then the minimum-volatility fully invested portfolio also has the maximum Sharpe ratio.
3. This result assumes that the portfolio returns are mean zero, which is a good approximation at a daily horizon.
4. This definition neglects the effect of changing weight due to daily price fluctuations. For the present exercise, however, such effects are small and can be safely ignored.
5. In this paper, we use the term alpha to mean the expected *total return* of the asset. In other contexts, alpha refers to the expected stock return in excess of the market component. Such a distinction is unimportant for present purposes.

References

Black, F. and Litterman, R. (1991). "Asset Allocation: Combining Investor Views with Market Equilibrium," *Journal of Fixed Income* **1**(2) (September), 7–18.

Ceria, S., Saxena, A., and Stubbs, R. (2012). "Factor Alignment Problems and Quantitative Portfolio Management," *Journal of Portfolio Management* **38**(2) (Winter), 29–43.

Grinold, R. and Kahn, R. (2000). *Active Portfolio Management* (Second Edition). McGraw-Hill.

Jorion, P. (1985). "International Portfolio Diversification with Estimation Risk," *Journal of Business* **58**(3), 259–278.

Kritzman, M. (2006). "Are Optimizers Error Maximizers?," *Journal of Portfolio Management* **32**(4) (Summer), 66–69.

Kritzman, M., Page, S., and Turkington, D. (2010). "In Defense of Optimization: The Fallacy of 1/N," *Financial Analysts Journal* **66**(2) (March/April), 31–39.

Ledoit, O. and Wolf, M. (2003). "Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection," *Journal of Empirical Finance* **10**(5), 603–621.

MacKinlay, C. (1995). "Multifactor Models Do not Explain Deviations from the CAPM," *Journal of Financial Economics* **38**(1), 3–28.

Markowitz, H. (1952). "Portfolio Selection," *Journal of Finance* **7**(1), 77–91.

Menchero, J. and Ji, L. (2016). "Multi-Asset Class Risk Model (MAC2)," *Bloomberg Research Note*.

Michaud, R. (1989). "The Markowitz Optimization Enigma: Is 'Optimized' Optimal?," *Financial Analysts Journal* **45**(1) (January/February), 31–42.

Rosenberg, B. (1974). "Extra-Market Component of Covariance in Security Returns," *Journal of Financial and Quantitative Analysis* **9**(2) (March), 263–274.

Scherer, B. (2015). *Portfolio Construction and Risk Budgeting* (Fifth Edition). Risk Books.

Keywords: Mean-variance optimization; portfolio analytics