



Research Report

Active management in more than one asset and the role of
correlations

Lee Overlay Partners
June 2000

Reference Number: lop-0600a

Active management in more than one asset and the role of correlations

Abstract

In this article we derive an expression for the information ratio of a portfolio of investment processes. Using this expression we show that correlations are a far less important contributor to overall performance than the performances of the individual investment processes. Hence we provide a theoretical underpinning to currency overlay (the separate specialist management of the currency exposure in a portfolio). More generally, we show that allocating a portfolio to a suite of specialists rather than one generalist is optimal, largely irrespective of correlations.

I. Introduction

We can measure the performance of an active investment process using the information ratio, defined as average excess return divided by residual risk. Such a performance measure is well documented, and important work includes that of Grinold and Kahn (1999) with useful one-dimensional results documented by Lee Overlay Partners (1999a).

When several processes are combined it is important to know not just how the individual one-dimensional pieces perform, but also how the portfolio of active processes perform. The central part of this article is the derivation of an expression for the information ratio of a number of investment processes as a function of the individual component parts. We analyse several examples where our expression can be employed to derive both existing and new results in active portfolio management.

In their seminal book, Grinold and Kahn (1999) provide valuable insight into this multi-process situation with their so-called “fundamental law of active management” and the concept of breadth. A proportion of this paper discusses the relationships between this work and our own. Important extensions covered in our work are that we are able to examine the situation where individual performances are different and that we can analyse the role of correlations and their impact on performance in combining active processes.

The specific practical situation of interest to us is international asset management where asset management and currency management constitute the overall process. Currency overlay is the separate specialist management of the currency component. We examine how much larger the specialist’s *currency* information ratio has to be, relative to the asset manager’s *currency* information ratio, in order to produce a higher total information ratio. We conclude an important result that in the absence of correlations any increase will produce a higher information ratio, and even when non-zero correlations are prevalent anything other than a negligible increase in currency information ratio will produce a higher total information ratio.

In Section II we give an outline of the framework within which we are working. In Section III we provide the main result of the article and give the expression for the multivariate information ratio as a function of various covariance matrices. In Section IV we detail the assumptions behind our expression. In Section V we give several examples of our expression. In Section VI we examine how correlations impact performance and provide more examples. In Section VII we relate our work to existing literature. In

Section IX we discuss further work, and in Section X we give a summary of our work and some conclusions.

II. The framework

We consider active management in a universe of many risky assets. At a single point in time we make forecasts of the asset returns, then create investment positions as a function of our forecasts using modern portfolio theory. The excess return of our portfolio will then depend on our positions and the magnitude of the asset returns. We consider the time series of excess returns as this investment process is repeated over a large number of time intervals. We are interested in the long run performance of our strategy, and measure that performance by information ratio defined as the average excess return divided by the standard deviation of excess return.

In this framework, the information ratio will depend on the volatility of the asset returns, the volatility of our forecasts, and the volatility of the errors in our forecasts, all of which will be interrelated by our ability to forecast. The information ratio will also depend on the correlations of the aforementioned variables.

III. An expression for the information ratio of a number of active management processes

The information ratio, denoted by IR, of a number of active investment processes is given by

$$IR = \text{tr}(\Omega \mathbf{V}^{-1} \Lambda) / (\text{tr}(\Omega \mathbf{V}^{-1} \Sigma \mathbf{V}^{-1}) + \text{tr}(\Omega \mathbf{V}^{-1} \Lambda)^2)^{1/2}.$$

where $\text{tr}(\cdot)$ denotes the trace of a matrix. The matrix $\mathbf{V}$ is the covariance matrix of the forecast errors and is entirely the choice of the forecaster. It can be considered the forecast of return variance conditional on the forecast of expected return. The matrix Σ is the covariance matrix of the actual returns and can be observed directly from empirical data. The matrix Ω is the covariance matrix of the expected return forecasts. This variable measures the extent to which the forecasts move through time, and in the multivariate setting it also measures the extent to which they move together. It is an observable variable. Λ is the diagonal matrix of aggressiveness factors of each process, it allows for the scenario when the forecasts of a model are consistently smaller (or larger) than the returns. The derivation of our expression appears in the Appendix.

IV. Assumptions

The assumptions behind our expression of information ratio are fairly benign. We have assumed that in the long run our forecasts are unbiased and normally distributed, and that our errors are independent of our forecasts. We have assumed that in the long run the average asset return is equal to the market, or benchmark, average return. That is to say that there exists no systematic abnormal return to the assets. We have assumed that the asset returns are normally distributed. Finally we have assumed that the investment positions are derived using mean-variance optimisation theory.

V. Examples

We illustrate the use of our expression of information ratio with a series of examples. The first of these examples considers the scenario of a single asset where we

establish the one-dimensional results of Lee Overlay Partners (1999a). We then progress to include many, but independent, assets each with identically performing investment processes. Here we show that our results obey Grinold and Kahn's fundamental law of active management. The next progression is to consider many independent assets but now we assume that the investment processes of each asset perform differently. In doing this we have provided the mathematical underpinning to currency overlay (i.e. separately managing currency and international assets so as to benefit from the added performance of two specialists instead of one). Following this we consider correlations and their effect on performance. We illustrate that even if the assets are correlated, separating the investment processes (and therefore ignoring the correlations) will yield better performance since it means that each can be attributed to specialists. This then provides the mathematical underpinning to the widespread practice of using a suite of specialist managers to manage components of a single portfolio. We conclude by incorporating assets with different volatilities and show that for international fixed income portfolios the case for a specialist currency manager is very strong.

In our examples we have assumed a monthly investment horizon but have annualised all of our results.

Example 1: A one-dimensional example

In Example 1 we analyse our expression for IR in a basic one-asset case and recover existing results of Grinold and Kahn (1999) and Lee Overlay Partners (1999a).

We begin our examples by introducing the well-known measure termed information coefficient, hereafter denoted by IC. It is formally defined as the correlation between the forecasts and the actual returns.

We are fortunate that IC relates to the variable in our expression for IR, and to show this we assume there to be a single asset so that the four matrices Ω , V , Σ , and Λ are represented by the scalars ω^2 , v^2 , σ^2 , and λ . We then show that the information coefficient is given as $IC = \lambda\omega/\sigma$ (see Appendix for proof). With a small amount of algebraic rearrangement our expression for IR can be written as

$$IR = IC / (1 + IC^2)^{1/2}.$$

This result matches that given in Lee Overlay Partners (1999a). Moreover, Grinold and Kahn make the assumption that IC^2 can be considered negligible on the appropriate grounds that in practice an IC will rarely be bigger than 0.3. By plugging that assumption into our equation we have that

$$IR = IC$$

which matches Grinold and Kahn's fundamental law of active management for unit breadth (i.e. one signal).

Example 2: A multidimensional example

In Example 2 we analyse our expression for IR when there are many assets and recover Grinold and Kahn's fundamental law of active management.

We consider a multidimensional example, but where each one-dimensional process is identical and independent. We use the results from the first example to

populate the matrices. Mathematically these assumptions are represented as $\Sigma = \sigma^2 \mathbf{I}$, $\Lambda = \lambda \mathbf{I}$ and hence $\Omega = \text{IC}^2 \sigma^2 / \lambda^2 \mathbf{I}$, and finally $\mathbf{V} = \sqrt{v} \mathbf{I}$, where $\mathbf{I}$ is the identity matrix.

Substituting these equations into our expression for IR we can show that

$$\text{IR} = \text{BR}^{1/2} \text{IC} / (1 + \text{IC}^2 \text{BR})^{1/2}.$$

where BR denotes breadth. Given that we have one forecast per asset in this example, breadth equals the number of assets (which is the dimension of the matrices). Again by making the assumption that IC^2 is negligible, we have arrived at Grinold and Kahn's full version of the fundamental law of active management which states that

$$\text{IR} = \text{BR}^{1/2} \text{IC}.$$

Observing from the one-dimensional example that we have a relationship between IR and IC, it is easy to show that $\text{IC} = \text{IR} / (1 - \text{IR}^2)^{1/2}$. We can then express the multivariate information ratio in terms of the individual ones thus

$$\text{IR} = \text{BR}^{1/2} \text{IR}_0 / (1 + \text{IR}_0^2 (\text{BR}-1))^{1/2}.$$

where IR_0 denotes the individual information ratios.

This equation measures the role of breadth in improving performance. We have a mathematical expression that shows that diversifying with many asset forecasts improves portfolio return for the same level of risk. This can be seen graphically in Chart 1 where we have plotted the level of achievable information ratio as breadth increases. For example, if a single process yields an information ratio of 0.5, then six such processes will yield a combined information ratio of 1.2.

Example 3: Different information ratios example

In Example 3 we analyse our expression for IR when information ratios differ for different assets. We show the intuitive result that the maximum portfolio performance will be attained by using the best performing individual managers. There are two parts to example 3, where forecast variance is dealt with in two different ways.

Example 3(i)

We now begin to examine the situation where, in a multidimensional setting, individual information ratios may differ from each other. The specific example of interest is in the separation of international asset management into asset management and currency management.

Let us then assume an identical scenario to the previous example, but now we assume different ICs so that Ω is now given by $\Omega = \text{diag}(IC_i^2)\sigma^2/\lambda^2$ where IC_i denotes the information coefficient of the i th asset process. This then gives us that

$$IR = (\sum IC_i^2)^{1/2} / (1 + \sum IC_i^2)^{1/2}$$

where the summations go from 1 to BR.

Clearly by setting the ICs equal to each other we recover our previous result that

$$IR = BR^{1/2} IC / (1 + IC^2 BR)^{1/2}.$$

It is useful to express the total information ratio as a function of the individual information ratios. By substituting the ICs for their respective values of IR, we immediately have that

$$IR = (\sum (IR_i^2 / (1 - IR_i^2)))^{1/2} / (1 + \sum (IR_i^2 / (1 - IR_i^2)))^{1/2}$$

where IR_i denotes the information ratio of the i th asset process. In the two dimensional case this unwieldy expression can be simplified to

$$IR = (IR_1^2 - 2 IR_1^2 IR_2^2 + IR_2^2)^{1/2} / (1 - IR_1^2 IR_2^2)^{1/2}.$$

This is a very important expression, for it is easy to show that for any positive value of IR_2 , IR is greater than IR_1 . Moreover, we can show that for any value of IR_1 and IR_2 , if a higher IR_2 can be generated then a higher IR will be generated.

Example 3(ii)

In the first part of our analysis of different information ratios we assumed that the variances of the two forecast errors were equal. But it is theoretically correct to have a lower forecast error variance for the asset with higher information ratio because we can explain more of the overall variance through our forecasts. We can consider this intuitively as investing more aggressively where we have insight. The theoretically correct value for V is $\Sigma - \Lambda \Omega \Lambda$, which using the definitions of Σ , Λ , and Ω in Example 3(i) yields $V = \text{diag} (1 - IC_i^2) \sigma^2$. Substituting these values into our expression gives us the improved information ratio given by

$$IR = \{ \sum_i IC_i^2 / (1 - IC_i^2) \} / \{ \sum_i IC_i^2 / (1 - IC_i^2)^2 + (\sum_i IC_i^2 / (1 - IC_i^2))^2 \}^{1/2}.$$

As can be seen the level of mathematical complexity has increased, and the expression for total information ratio as a function of individual information ratios is given by the messy formula

$$IR = \{ \sum_i IR_i^2 / (1 - 2IR_i^2) \} / \{ \sum_i IR_i^2 (1 - IR_i^2) / (1 - 2IR_i^2)^2 + (\sum_i IR_i^2 / (1 - 2IR_i^2))^2 \}^{1/2}.$$

Even in the two-dimensional case we cannot simplify it to any particularly elegant form.

One alternative is

$$IR = (IR_1^2 + IR_2^2 - 4 IR_1^2 IR_2^2) / (IR_1^2 + IR_2^2 - 6 IR_1^2 IR_2^2 + 8 IR_1^4 IR_2^4)^{1/2}.$$

More important is the fact that for any value of IR_1 and IR_2 , if a higher IR_2 can be generated then a higher IR will be generated.

Here we have derived the mathematical reasoning behind the case for currency overlay. *Under the assumption that a specialist currency manager will produce a higher information ratio on currencies than the international asset manager will produce on currencies, the total information ratio of the international portfolio will be higher by using a currency overlay manager.*

The increase in performance is not trivial. This can be clearly seen in Chart 2 where we plot the total information ratio as a function of varying currency management information ratios. We have highlighted two points on Chart 2. Point A represents typical numbers that one might expect from an international equity manager: an asset information ratio of about 0.75 and a currency information ratio of about 0.2 providing a total information ratio of 0.77. Point B represents the same equity information ratio, but includes currency overlay that will typically be yielding an information ratio of about 1.0. This gives a total information ratio of 1.21, almost doubling overall performance.

Also plotted on Chart 2 is the information ratio computed without adjusting V to account for the level of IC (i.e using the formula in the Example 3(i)). The fact that the two lines are indistinguishable is evidence that there is no significant difference in performance from adjusting V . In fact, paradoxically there is a slight reduction in performance by adjusting V . This can be explained by the fact that mean-variance optimisation, which we used to derive our investment positions, does not maximise the

information ratio. Lee Overlay Partners (1999a) discusses this for the one-dimensional case where a mean-second moment approach is shown to be optimal. In the Appendix we extend this proof to the multi-dimensional case (although it is beyond the remit of this article to derive an expression for the information ratio from using the optimal investment approach). We know that the second moment must be at least as large as the variance of a distribution, and therefore the reduction in the size of $\mathbf{V}$ that we made to account for the better performance actually produced a less optimally constructed portfolio. However, we reiterate that the difference is negligible.

VI. Incorporating correlations

We now take our analysis to the next level of complexity by considering correlations. There are three correlation matrices in our expression for information ratio. They are the correlation matrices embedded in the variables $\mathbf{V}$, $\mathbf{\Omega}$, and $\mathbf{\Sigma}$.

Recall that $\mathbf{V}$ is the covariance matrix of the forecast return conditional on the forecast of expected return. In this capacity it controls the extent to which the expected return forecast of one asset can influence the investment position for another asset. The correlation of $\mathbf{V}$ is the only mechanism for doing this. As an example, the only way to ensure that the investment position for an asset depends only on that asset's expected return, is to set the correlations of $\mathbf{V}$ equal to zero.

The correlations in $\mathbf{\Sigma}$ are the actual observable correlations of the asset returns. When we think of a correlation between two assets, it is the correlation of $\mathbf{\Sigma}$ that we are typically considering.

The final correlation is that of Ω and is perhaps the least understood of all the inputs. It is the correlation between the expected return forecasts. There is often confusion between the correlation of Ω and the correlation of V . The difference is as follows. *For a single forecast* the correlation of V measures how much the forecast error of one asset relates to the forecast error of another (by forecast error we mean the difference between the actual return and our forecast). Whereas Ω measures how much the change in a forecast *through time* relates to the change in another forecast.

In relating Ω , V , and Σ to each other, we consider the actual return to be equal to the forecast of expected return plus the forecast error. The covariance matrix of the first term is Ω and we shall denote the covariance of the error term to be Ψ noting that we can consider V our estimate of Ψ (in fact V is set equal to Ψ in the work of Grinold and Kahn). When the errors and forecasts are independent the matrices are related by the equation $\Sigma = \Omega + \Psi$ (for clarity of discussion we have taken $\Lambda = \mathbf{I}$). *Therefore an observed asset return correlation can be attributed to **both** the correlation of the forecast **and** the correlation of the error.* In viewing our process in this fashion we can clarify a deeply misunderstood issue regarding the role of correlations in practical investment management.

In practice, the temptation is to account for the observed correlation by setting the correlation of the forecast errors (which is the correlation in our matrix V) equal to it. But this implies that all of the observed correlation is attributed to the errors in the forecasts and none to the forecasts themselves. Is it not equally plausible to attribute the observed correlation to the correlation of the forecasts? At least in the latter case when two assets have a high correlation their two forecasts are similar. In fact it could be argued that all of

the observed correlation should be attributed to the forecasts since this implies that the error we may make in one forecast is independent of the error in another. In such a scenario the correlation of $\mathbf{V}$ should be set to zero *even though the observed correlation of Σ is non-zero*. Moreover it would be erroneous to set the correlation of $\mathbf{V}$ to be anything other than zero.

There is an appeal in constructing a set of models whose forecasts move together when assets are known to be correlated and whose investment positions are based solely on the forecast for that asset. This construction requires that the correlations of Σ are accounted for in our forecast covariance matrix Ω , and that we use a zero correlation in $\mathbf{V}$. It is further discussed in Lee Overlay Partners (1999b).

Example 4

In Example 4 we analyse our expression for information ratio with respect to correlations. We show that correlations do not significant impact overall information ratio, and the search for higher individual information ratios is more important than the use of correlations.

Example 4 is split into three parts. The first part considers all asset processes to be identical and we show that correlations act as an adjustment to breadth. The second part considers differing information ratios and we show that ignoring correlations in order to use higher information ratios on the individual components is optimal. The third part is a case study of correlations in the management of currency in an international fixed income portfolio. We conclude that there is a case for currency overlay on international fixed income portfolios.

Example 4(i)

To focus on correlations we take the scenario detailed in Example 2 where the individual processes were identical. But now we assume that these processes are not independent and instead have some non-zero correlation. We represent this algebraically as $V = \sqrt{v}^2 \Phi$, $\Sigma = \sigma^2 \Pi$, and $\Omega = \omega^2 \Theta$ where Φ , Π , and Θ are correlation matrices.

Substituting these variables into our expression for information ratio we have that

$$IR = k^{1/2} IC / (1 + IC^2 k)^{1/2}$$

where $k = \text{tr}(\Theta \Phi^{-1})^2 / \text{tr}(\Theta \Phi^{-1} \Pi \Phi^{-1})$. Clearly this expression is identical to our expression for information ratio in Example 2 except that k now replaces BR. We can thus consider the impact of correlations to be similar to adjusting breadth.

We can further analyse this adjusting variable, k , by considering the two-dimensional scenario. By inverting Φ and expanding the traces, we have that

$$k = BR (1 - \phi\theta)^2 / ((1 - \phi\theta)(1 - \phi\pi) + (\theta - \phi)(\pi - \phi))$$

where the correlations are now scalars, and BR equals 2. A contour plot showing how k is affected by θ and ϕ , for a given π , is shown in Chart 3.

The values of θ and ϕ that maximise k for a given value of π are $\phi = \pi$ and $\theta = -\text{sgn}(\pi)$ (see Appendix for a derivation). Of interest here is the fact that the optimal settings depend only on π and that $\phi = \pi$ is optimal regardless of the value of θ . More interesting again is that the information ratio will be maximised by making the expected return forecasts as correlated as possible, either positively or negatively, irrespective of the actual correlations. And these huge correlations should be in the opposite direction to the actual correlations. These characteristics of ϕ and θ are clearly seen in Chart 3.

We are left with an interesting decision, for the information ratio is maximised by using a correlation for θ that is totally different to the actual returns correlations. Would it not be more appropriate, in a theoretical sense, to use correlations that correspond to the actual correlation? For example, we have assumed that $\Sigma = \Lambda\Omega\Lambda + \mathbf{V}$ and we thus have the relationship that $\pi = IC^2\theta + (1-IC^2)\phi$. Whilst it may be more optimal to use extreme values for θ , we cannot avoid the fact that θ exists as a relationship between two forecasting models, and we therefore proceed by incorporating our equation relating π , θ , and ϕ . Optimising with the additional constraint gives us that the optimal value of θ is π .

Substituting the optimal values of θ and ϕ into our expression for k , we have that $k = BR$. Incidentally, substituting the $\theta = 0$ and $\phi = 0$ also gives us that $k = BR$.

These are interesting results. The best we can achieve by incorporating correlations is the same as the scenario when everything was independent. Moreover, irrespective of the actual correlation, we can do no better than taking both θ and ϕ equal to zero. Put another way, it is entirely possible to reduce performance by trying to account for correlations in an investment process, even if the estimates of those correlations are good.

A scenario more related to practice is one where we do have a correlation in our modelling process so that $\theta = \pi$, but for portfolio construction purposes we wish to keep our investment positions a function solely of each asset return forecast so that $\phi = 0$. Then $k = BR / (1 + \pi^2)$. Immediately we can see that even for moderate values of π , the change in k is very small. For example, a correlation of 0.3 yields a reduction in k of only 8%. The corresponding change to information ratio is also very small. This is seen in Chart 4 where the information ratio is plotted for various values of π with a fixed IC of 0.3.

Highlighted on the chart is the previous example where $k = 0.3$. The reduction in information ratio can be seen to be 0.08.

The fact that this change in performance from ignoring correlations is small is crucial to the argument that allocating component parts of a portfolio to investment specialists (and therefore ignoring correlations) is an optimal approach. But to complete the case we must examine the scenario where correlations are present but individual information ratios differ for each component part.

Example 4(ii)

We consider the above scenario but now use different information ratios. Again we view the two dimensional example in the context of an international asset portfolio plus the currency portfolio.

Let us suppose that an asset specialist manages the asset component of the international portfolio and consider two cases. First, the case where the asset manager also manages the currency portfolio and accounts for the forecast error correlations in the portfolio construction. Second, the case where the asset manages the assets and a currency specialist manages the currencies, but they do so without accounting for each other and hence assume a forecast error correlation of zero.

We can readily incorporate differing information ratios, and it is easy to show that

$$IR = k^{1/2} IC / (1 + IC^2 k)^{1/2}$$

but now k is given by

$$k = BR \gamma (\delta - \phi\theta)^2 / ((\delta - \phi\theta)(1 - \phi\pi) + (\theta - \delta\phi)(\pi - \phi))$$

where γ is the ratio of the two information coefficients, and $\delta = (1 + \gamma^2)/2\gamma$. This expression has an extra variable to the one in the previous example, but it is pleasing to note that it is still maximised by $\phi = \pi$ (see Appendix for details).

We continue with our previous assumption that $\theta = \pi$ and consider two scenarios. The first is where the asset manager is managing both the assets and the currencies, and does so by incorporating their correlation by setting $\phi = \pi$. In such a scenario the corresponding k , denoted by k_A , is given as

$$k_A = \text{BR } \gamma_A (\delta_A - \pi^2) / (1 - \pi^2).$$

In the second scenario we consider that a specialist is now managing the currency component of the international portfolio, and hence the correlation is not being incorporated thus $\phi = 0$. In this second scenario the corresponding k , denoted by k_C , is given as

$$k_C = \text{BR } \gamma_C \delta_C^2 / (\delta_C + \pi^2).$$

Obviously the reason for using a specialist is that we expect the information ratio to be higher, that is we expect γ_C to be higher than γ_A . But we wish to examine how much lower the *currency* information ratio of the asset manager can be relative to the currency manager yet still match total information ratio by incorporating the correlations. While heavily mathematical, a solution to this question does exist and it can be shown that the asset manager's currency information ratio, denoted by IR_A , must be at least as big as

$$\text{IR}_A = x / (1 + x^2)^{1/2}$$

where $x = [\pi^2 + (\pi^4 + 2 \gamma_C^2 \delta_C^2 (1 - \pi^2) / (\delta_C + \pi^2) - 1)^{1/2}] \text{IR}_0 / (1 - \text{IR}_0^2)^{1/2}$ with IR_0 denoting the asset manager's asset only information ratio, and $\gamma_C = [\text{IR}_C / (1 - \text{IR}_C^2)^{1/2}] / [\text{IR}_0 / (1 - \text{IR}_0^2)^{1/2}]$ where IR_C denotes the specialist's currency information ratio.

Clearly this expression does not lend itself to an intuitive examination. Instead, we graphically represent it in an example in Chart 5. We continue our ongoing example and use the numbers of Example 3(ii) where the asset only information ratio is 0.75 and the currency overlay information ratio is 1.0. In Chart 5 we plot the level of *currency* information ratio that is required by the asset manager to match that of the specialist. We observe that even when the correlation is as high as 0.5, the asset manager's *currency* information ratio must be within 85% of the specialist's. *We conclude that correlations are not an issue in allocating to a specialist.*

Intriguingly there is another value of IR_A that will match k_A and k_C , and is given by $x = [\pi^2 - (\pi^4 + 2 \gamma_C^2 \delta_C^2 (1 - \pi^2) / (\delta_C + \pi^2) - 1)^{1/2}] IR_0 / (1 - IR_0^2)^{1/2}$. But this is normally negative and therefore unfeasible since volatility must be positive. However, if IR_C is very small this alternative value of IR_A can be positive. This occurs because the information ratio is so small that we are getting a bigger impact from the role of correlations. This is something of an academic argument because the increase in total information ratio at this very low level of individual information ratio is very small. In Chart 6 we show the total information ratio for a range of individual information ratios and this correlation effect at the low performance end can be seen. With a correlation of 0.3, and an individual information ratio of 0.75, once the information ratio of the second component has gone beyond 0.14 the role of correlations has vanished. And even between 0 and 0.14, the reduction in information ratio is no greater than about 0.001.

Example 4(iii)

Our final example is an extension of the previous scenario but incorporating different asset volatilities. We examine the case for currency overlay on international fixed income portfolios. Here we have an example of our expression where the volatility of the asset returns is considerably less than the volatility of currency. This is a particularly important example in our review of correlations because the case for currency overlay on international fixed income has been argued to be inappropriate on the grounds of non-zero correlations.

We begin by populating our matrices with appropriate value. Typical asset assumptions for bonds and currencies are that the volatilities are 5% and 10% respectively on an annualised basis, and that their correlation is 0.2 (actually this number is high relative to empirical data (1978 – 1999) suggesting a correlation of 0.13). These make up our covariance matrix Σ .

We shall assume that the expected return forecast correlation is equal to the asset return correlation and thus equals 0.2.

A good international fixed income manager should attain an information ratio of 0.5 for hedged fixed income management, and 0.3 for currency management. We assume that the fixed income manager uses the optimal forecast error correlation of 0.2 in the portfolio construction. These inputs lead to a total information ratio of 0.58.

We now consider the situation where the currency component is being managed independently by a specialist. The forecast error correlation must then be 0, but the information ratio for the currency component is now 1.0. This leads to a total information ratio of 1.08. This is very significant increase in performance and can only be matched by

the fixed income manager producing a currency information ratio of 0.98. Chart 7 plots the increase in total information ratio. *The case for currency overlay on international fixed income portfolios is clearly strong.*

VII. Relating our results to Grinold and Kahn

Our structure of analysis is very similar to that of Grinold and Kahn (1999), and so our results are similar. In fact in our first two examples, we considered two special cases when our expression matches Grinold and Kahn's fundamental law of active management.

There are however several differences between the two bodies of work and we highlight those here.

Grinold and Kahn consider a set of signals and relate them to actual returns via a correlation matrix. They then derive an expression for the forecasts of expected returns. We differ in that we begin by considering the forecasts directly but note that we can reverse Grinold and Kahn's derivation to generate their signals.

In Grinold and Kahn's work, the relationship between the forecast and the actual return is that the actual return equals the forecast plus an independent error term (i.e. $r = f + \varepsilon$ where r denotes the actual return, f denoted the forecast, and ε denotes the error). We include an aggressiveness, or scaling, term so that the actual return is equal to a scaled version of the forecast plus an independent error term (i.e. $r = \lambda f + \varepsilon$ where λ denotes the scaling factor). This is an important difference between the two approaches in that it allows us to incorporate any level of forecast variance, whereas in the Grinold and Kahn model the forecast variance is deterministically set as a function of forecast performance.

Since both we and Grinold and Kahn use a mean-variance approach to deriving investment positions, the variance of the forecast becomes a key influence on the magnitude of those investment positions. Our scaling parameter therefore allows us to control the aggressiveness of these positions.

The impact of this difference in modelling is best illustrated in an example. Suppose an investment manager has a low skill level in making forecasts then according to the Grinold and Kahn model they must (with mathematical certainty) be taking small investment positions. Our framework allows us to account for those poor quality managers that continue to take aggressive investment decisions. Conversely, our model also allows us to account for the managers who are very good forecasters but habitually take small positions. Indeed it may be part of the remit given to the active manager that a specific level of aggressiveness is targeted. Moreover, irrespective of an ability to forecast, an issue with the omission of this scaling factor is that an upper bound for forecast variance is the actual return variance and that bound is only attained with perfect foresight. It is therefore impossible to incorporate the situation where forecast variance is greater than asset variance.

Having discussed the importance of a scaling factor we point out that a forecasting approach based in formal quantitative methods will yield a scaling factor of 1 and a variance equal to its mathematical derivation. Our results then become identical to those of Grinold and Kahn.

In addition to scaling there is another benefit of allowing the forecast variance to take any value as opposed to being imposed mathematically. When there is more than one asset being modelled, correlations between forecasts must be provided. By

investment design these correlations may be different from the mathematically derived numbers. For example if international assets are modelled they may be broken up into the assets plus the currencies. Each part may be given to different managers and therefore their investment positions will be independent of each other. This is characterised by a forecast correlation of zero. This kind of scenario cannot be accounted for in the Grinold and Kahn approach.

There is another, more mathematical difference between Grinold and Kahn's derivation of the Fundamental Law of Active Management and our own result. It is worthy of some discussion but its mathematical nature warrants deference to the Appendix.

VIII. Further Work

We have examined several scenarios in this article, all using our expression for information ratio $IR = \text{tr}(\Omega \mathbf{V}^{-1} \Lambda) / (\text{tr}(\Omega \mathbf{V}^{-1} \Sigma \mathbf{V}^{-1}) + \text{tr}(\Omega \mathbf{V}^{-1} \Lambda)^2)^{1/2}$. Owing to the quantity of inputs, however, there are many other examples that can be analysed. It is simply unfeasible to incorporate all of these into one body of work. In particular we have neglected to analyse examples where Ω is not related to Σ by IC. Such an example exists when the process forecast volatility is not a function of performance, but instead the process targets an average return level. Targeting is not uncommon in practice where ICs and IRs cannot be guaranteed and instead a risk control approach is used. If average return targets are set then we can use the equation $\alpha_0 v^2 = \omega^2 \lambda$, where α_0 is the target, in order to derive values for v , ω , or λ . Such a construction is perhaps the next step of research.

We should also consider relaxing some of the assumptions relating Σ , Ω , Λ , Ψ , and $\mathbf{V}$ via the equations $\Sigma = \Lambda\Omega\Lambda + \Psi$ and $\mathbf{V} = \Psi$. It is certainly the case that out-of-sample we should not expect such relationships to hold and the subsequent impact on performance would be interesting to analyse.

Correlations are notoriously difficult to forecast, yet all the results we have seen have assumed that they are known. This leads to an unknown impact of correlation estimation error. Further work is clearly required to examine this, and we suspect that the result will act to strengthen the argument that assuming a zero correlation is an appropriate approach. Hence the case for specialists is made even more concrete.

IX. Summary and Conclusions

The key result of this article is that we have derived an expression for portfolio information ratio as a function of risks and correlations of actual returns, forecast returns, and forecast errors. This expression can be used to evaluate the performance of any portfolio of investment processes. We have related our result to existing literature, but more importantly we have argued the case for currency overlay and more broadly the general case for specialist managers. We have shown that with, or without, non-zero correlations the case for specialist management is equally strong, since superior individual information ratio is dominant over any impact from correlations.

Appendix

A.I. A derivation of our formula

Suppose that *the distribution of expected return forecasts* denoted by the vector $\mathbf{m}$ is given as

$$\mathbf{m} \sim N_p(\mathbf{0}, \Omega)$$

where p is the dimension of $\mathbf{m}$, and $\mathbf{0}$ is a vector of zeros.

At a point in time, our forecast of the return distribution conditional on our expected return forecast, denoted by $\mathbf{f}$, is

$$\mathbf{f}|\mathbf{m} \sim N_p(\mathbf{m}, \mathbf{V})$$

where $\mathbf{V}$ is the covariance of the distribution conditional on our forecast of expected return, and is assumed to be known.

The actual return vector conditional on the expected return forecast, denoted by $\mathbf{r}$, is proportional to the forecast plus an independent error term. We write that *at a point in time, the actual return distribution conditional on our expected return forecast is*

$$\mathbf{r}|\mathbf{m} \sim N_p(\Lambda\mathbf{m}, \Psi).$$

The error term measures the quality of the forecasting. When the forecasting is good the variance of the error is small, but when the forecast is poor the variance of the error is larger. The matrix Λ is a diagonal matrix to account for the situation where the variance of the forecast of expected returns is not consistent with the variance of the actual expected returns. For many statistical approaches to deriving forecasts we can take Λ to equal the identity matrix denoted by $\mathbf{I}$.

Modern Portfolio Theory (see Bodie, Kane, and Marcus 1986 for example) suggests that the optimal investment positions are given by maximising the expected portfolio return subject to the portfolio variance (although see Lee Overlay Partners 1999a for a refute of this theory when expected return forecasts are not constant through time, e.g. $\Omega \neq 0$). We have assumed that the risk free alternative is open to the investor, and that there are no investment constraints. We therefore do not have to ensure our positions sum to a specific value as the risk free asset will take up the slack in that constraint. We therefore have that the optimal investment positions, denoted by the vector y , are given as

$$y = V^{-1}m.$$

The return on our portfolio at a point in time, denoted by the scalar x , is the product of the positions and the actual returns and therefore given by

$$x = y'r = m'V^{-1}r$$

and so the portfolio return distribution, conditional on our forecast of expected return, is given as

$$x \sim N(m'V^{-1}\Lambda m, m'V^{-1}\Psi V^{-1}m).$$

As a measure of performance we are interested in the information ratio, denoted by IR, and defined as

$$IR = E[x] / \text{Stdev}[x].$$

We express the information ratio in the form used in Lee Overlay Partners (1999a) thus

$$IR = (E[x^2] / E[x]^2 - 1)^{-1/2}$$

and are left to derive $E[x]$ and $E[x^2]$. In order to do this we follow the reasoning laid out in Lee Overlay Partners (1999a) noting that the marginal expectations of x and x^2 are given as

$$E[x] = E_z[E[x|z]] \text{ and } E[x^2] = E_z[E[x^2|z]].$$

The first two zero centred moments of x which are extracted directly from the distribution of x thus

$$E[x|z] = \mathbf{m}'\mathbf{V}^{-1}\mathbf{\Lambda}\mathbf{m} \text{ and } E[x^2|z] = \mathbf{m}'\mathbf{V}^{-1}\mathbf{\Psi}\mathbf{V}^{-1}\mathbf{m} + (\mathbf{m}'\mathbf{V}^{-1}\mathbf{\Lambda}\mathbf{m})^2.$$

All that remains is to derive the marginal expectations of x and x^2 from our conditional distributions.

The first of these expressions is easily shown to be

$$E[x] = \text{tr}(\mathbf{\Omega}\mathbf{V}^{-1}\mathbf{\Lambda}),$$

and using the result in Lee Overlay Partners (2000) the second expression is given by

$$E[x^2] = \text{tr}(\mathbf{\Omega}\mathbf{V}^{-1}\mathbf{\Psi}\mathbf{V}^{-1}) + \text{tr}(\mathbf{\Omega}\mathbf{V}^{-1}\mathbf{\Lambda}\mathbf{\Omega}\mathbf{V}^{-1}\mathbf{\Lambda}) + \text{tr}(\mathbf{\Omega}\mathbf{V}^{-1}\mathbf{\Lambda}\mathbf{\Omega}\mathbf{\Lambda}\mathbf{V}^{-1}) + \text{tr}^2(\mathbf{\Omega}\mathbf{V}^{-1}\mathbf{\Lambda}).$$

These two expressions can be plugged into the equation for information ratio so we have that

$$\text{IR} = \text{tr}(\mathbf{\Omega}\mathbf{V}^{-1}\mathbf{\Lambda}) / (\text{tr}(\mathbf{\Omega}\mathbf{V}^{-1}\mathbf{\Psi}\mathbf{V}^{-1}) + \text{tr}(\mathbf{\Omega}\mathbf{V}^{-1}\mathbf{\Lambda}\mathbf{\Omega}\mathbf{V}^{-1}\mathbf{\Lambda}) + \text{tr}(\mathbf{\Omega}\mathbf{V}^{-1}\mathbf{\Lambda}\mathbf{\Omega}\mathbf{\Lambda}\mathbf{V}^{-1}))^{1/2}.$$

This expression can be tidied up by noting that the marginal variance of $\mathbf{r}$, denoted by $\mathbf{\Sigma}$, is defined as $\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Omega}\mathbf{\Lambda} + \mathbf{\Psi}$ thus

$$\text{IR} = \text{tr}(\mathbf{\Omega}\mathbf{V}^{-1}\mathbf{\Lambda}) / (\text{tr}(\mathbf{\Omega}\mathbf{V}^{-1}\mathbf{\Sigma}\mathbf{V}^{-1}) + \text{tr}(\mathbf{\Omega}\mathbf{V}^{-1}\mathbf{\Lambda})^2)^{1/2}.$$

A.II. An expression for the information coefficient

The information coefficient is defined as the correlation between returns and forecasts. Using our notation, that is

$$IC = (E[rm] - E[r]E[m]) / (V[r]V[m])^{1/2}.$$

Now, we know by definition that $r = \lambda m + \varepsilon$, and that $E[m]=0$ therefore

$$IC = E[\lambda m^2] / \sigma \omega$$

where σ is the standard deviation of r , and ω is the standard deviation of m . By noting that $E[\lambda m^2] = \lambda \omega^2$, we have the desired result that

$$IC = \lambda \omega / \sigma.$$

A.III. Optimal investment strategies in the multidimensional case

This derivation is a multivariate extension of the univariate version in Lee Overlay Partners (1999a).

The objective is to find the optimal investment position vector, denoted by $\mathbf{a}_t$ as a function of return forecast vector, $\mathbf{m}_t$, which maximises the information ratio. We begin by noting that maximising the information ratio is equivalent to minimising the function, h , defined as

$$h = E[\mathbf{a}_t^T \mathbf{r}_t \mathbf{r}_t^T \mathbf{a}_t] / E[\mathbf{a}_t^T \mathbf{r}_t]^2 = E_m[E[\mathbf{a}_t^T \mathbf{r}_t \mathbf{r}_t^T \mathbf{a}_t | \mathbf{m}_t]] / E_m[E[\mathbf{a}_t^T \mathbf{r}_t | \mathbf{m}_t]]^2.$$

Given that $\mathbf{a}_t$ is a function of $\mathbf{m}_t$, the conditional expectations can be written as

$$E[\mathbf{a}_t^T \mathbf{r}_t | \mathbf{m}_t] = \mathbf{a}_t^T E[\mathbf{r}_t] = \mathbf{a}_t^T \Lambda \mathbf{m}_t, \text{ and } E[\mathbf{a}_t^T \mathbf{r}_t \mathbf{r}_t^T \mathbf{a}_t | \mathbf{m}_t] = \mathbf{a}_t^T E[\mathbf{r}_t \mathbf{r}_t^T] \mathbf{a}_t = \mathbf{a}_t^T (\Lambda \mathbf{m}_t \mathbf{m}_t^T \Lambda + \Psi) \mathbf{a}_t.$$

We can therefore write h as

$$h = E[\mathbf{a}_t^T (\Lambda \mathbf{m}_t \mathbf{m}_t^T \Lambda + \Psi) \mathbf{a}_t] / E[\mathbf{a}_t^T \Lambda \mathbf{m}_t]^2.$$

At this point we need to use a multivariate version of the Schwarz inequality to obtain an upper bound for h . In order to do this we follow the logic of Lindgren (1976). We consider the non-negative random variable $(\mathbf{U} - c\mathbf{V})^T(\mathbf{U} - c\mathbf{V})$ where $\mathbf{U}$ and $\mathbf{V}$ are multivariate random variables and c is a scalar constant. The expectation of this expression must be greater, or equal to zero, and by expanding the quadratic this gives us that

$$E[\mathbf{U}^T \mathbf{U}] + c^2 E[\mathbf{V}^T \mathbf{V}] - 2c E[\mathbf{U}^T \mathbf{V}] \geq 0.$$

For this to be so for all c , the discriminant must be non-positive, and so

$$E[\mathbf{U}^T \mathbf{V}]^2 - E[\mathbf{U}^T \mathbf{U}] E[\mathbf{V}^T \mathbf{V}] \leq 0$$

which, when rearranged in a form that suits our analysis, gives us that

$$E[\mathbf{U}^T \mathbf{U}] / E[\mathbf{U}^T \mathbf{U}]^2 \geq E[\mathbf{V}^T \mathbf{V}]$$

with equality holding only if $\mathbf{U} = c\mathbf{V}$ for some c .

We set $\mathbf{U} = \mathbf{L}_t^T \mathbf{a}_t$ and $\mathbf{V} = \mathbf{L}_t^{-1} \Lambda \mathbf{m}_t$ where $\mathbf{L}_t$ is the Cholesky decomposition of $\Lambda \mathbf{m}_t \mathbf{m}_t^T \Lambda + \Psi$ so that $\mathbf{L}_t \mathbf{L}_t^T = \Lambda \mathbf{m}_t \mathbf{m}_t^T \Lambda + \Psi$. $\mathbf{L}_t$ must exist since Ψ is a covariance matrix and $\Lambda \mathbf{m}_t \mathbf{m}_t^T \Lambda$ can trivially be shown to be positive definite (see Rencher 1998 for details of Cholesky decompositions). Using our version of Schwarz's inequality, we have that

$$h \geq 1 / E[\mathbf{m}_t^T \Lambda (\Lambda \mathbf{m}_t \mathbf{m}_t^T \Lambda + \Psi)^{-1} \Lambda \mathbf{m}_t].$$

We know that equality only holds if $\mathbf{U} = c\mathbf{V}$, equivalently if

$$\mathbf{a}_t = c (\Lambda \mathbf{m}_t \mathbf{m}_t^T \Lambda + \Psi)^{-1} \Lambda \mathbf{m}_t.$$

If we assume a mainstream statistical approach to our modelling so that $\Lambda = \mathbf{I}$ and $\Psi = \mathbf{V}$, then the optimal position vector is given as $\mathbf{a}_t = c(\mathbf{m}_t \mathbf{m}_t^T + \mathbf{V})^{-1} \mathbf{m}_t$ which is simply the result of a mean-second moment optimisation, and *not* mean-variance optimisation.

A.IV. The value of ϕ and θ that maximises k

The variable k measures how much the information ratio is affected by correlations and is considered an adjustment to breadth. It is given as

$$k = \text{BR } \gamma (\delta - \phi\theta)^2 / ((\delta - \phi\theta)(1 - \phi\pi) + (\theta - \delta\phi)(\pi - \phi))$$

where γ is the ratio of the two information coefficients, and $\delta = (1 + \gamma^2)/2\gamma$.

Solving for ϕ :

By differentiating k with respect to ϕ and setting equal to zero, we have that

$$((\delta - \phi\theta)(1 - \phi\pi) + (\theta - \delta\phi)(\pi - \phi))\theta - (\delta - \phi\theta)(\pi\delta - \phi\theta\pi - \delta\phi + \theta) = 0$$

which can be rearranged to

$$(\pi - \phi)(\theta^2 - \delta^2) = 0$$

and hence k is maximised by setting $\phi = \pi$.

Solving for θ :

As a function of θ , we observe that k can be written as

$$k = \gamma (\alpha - \theta)^2 / (\theta + \beta)$$

where α , β , and γ are constants (i.e. not including θ). By differentiating k with respect to θ we have that

$$dk/d\theta = -2\gamma z - \gamma z^2$$

where $z = (\alpha - \theta) / (\theta + \beta)$. Taking the second derivative is thus given as

$$d^2 k/d\theta^2 = -2\gamma (1 + z) dz/d\theta,$$

and by noting that $dz/d\theta = -(\alpha + \beta) / (\theta + \beta)^2$ and that $1 + z = (\alpha + \beta) / (\theta + \beta)$, we have that

$$d^2k/d\theta^2 = 2\gamma ((\alpha + \beta) / (\theta + \beta))^2 / (\theta + \beta)$$

which can be rewritten as

$$d^2k/d\theta^2 = 2 \{(\alpha + \beta) / ((\theta + \beta)(\alpha - \theta))\}^2 k.$$

Since k is always positive, and the only other term in $d^2k/d\theta^2$ is a squared term, the second derivative must always be positive, therefore there can be no local maximum as a function of θ . The maximum value of k must then be given at one of the end points $\theta = \pm 1$. When $\phi = \pi$, $k = (\phi - \pi\theta) / (1 - \pi^2)$ and this can clearly be seen to be higher when $\theta = -\text{sgn}(\pi)$.

A.V. Comparing the mathematics of Grinold and Kahn

Both approaches define information ratio to be

$$IR = E[\alpha] / \sqrt{V[\alpha]}$$

where α is the active return from the investment process. Both approaches proceed by showing that

$$IR = E[E[\alpha|f]] / \sqrt{E[V[\alpha|f]]}.$$

But Grinold and Kahn then make the approximation that

$$E[E[\alpha|f]] / \sqrt{E[V[\alpha|f]]} = \sqrt{E[E[\alpha|f]^2] / V[\alpha|f]}.$$

Although making the ensuing mathematics a great deal simpler, such an approximation does require a fairly in-depth statistical analysis to measure its validity. The reader is encouraged to review the theory surrounding Jensen's inequality (see Lindgren 1976) in order to do this.

In our work, we have derived $E[\alpha]$ and $V[\alpha]$ directly.

References

- Bodie, Z., Kane, A., and Marcus, A. J., (1986) *Investments*. Third edition. Irwin.
- Lee Overlay Partners, (1999a) Naïve and optimised investment strategies: a paradox, *Lee Overlay Partners Research Report. Number: lop-0799a*.
- Lee Overlay Partners, (1999b) A discussion on allocation rules, and a recommended approach for use in currency overlay, *Lee Overlay Partners Research Report. Number lop-0799b*.
- Lee Overlay Partners, (2000) The fourth moment of a multivariate normal distribution, *Lee Overlay Partners Research Report. Number lop-0300*.
- Grinold, R. C. and Kahn, R. N., (1999) *Active portfolio management*. Second Edition. McGraw-Hill.
- Lindgren, B. W., (1976) *Statistical Theory*. Third Edition. Macmillan Publishing Co., Inc.
- Rencher, A. C. (1998) *Multivariate statistical inference and applications*. John Wiley and Sons.

Chart 1. Caption

Increasing breadth, the number of signals, improves the information ratio through diversification of investments.

Chart 2. Caption

The total information ratio of an international asset portfolio will be improved by a better currency information ratio. With typical international asset management numbers such as asset information ratio of 0.7 and currency information ratio of 0.2, using a currency specialist with a currency information ratio of 1.0 will double the total information ratio.

Chart 3. Caption

The information ratio of two equal processes will be maximised by using a forecast error correlation equal to the actual return correlation. The information ratio is also maximised by taking the forecast correlation equal to ± 1 .

Chart 4. Caption

The total information ratio is negligibly impacted by even moderate sizes of correlation. A correlation of 0.3 reduces total information ratio by about 8% when IC equals 0.3.

Chart 5. Caption

Regardless of correlation, the asset manager must provide almost the same currency information ratio as the specialist to match the overall performance of the asset manager plus the currency specialist.

Chart 6. Caption

The total information ratio will generally increase as the individual information ratios increase, regardless of correlations. At very low levels of individual information ratios, the correlations will have an impact and a reduction in total information ratio can occur.

Chart 7. Caption

International fixed income portfolios are better served by using a currency specialist. With typical international asset management numbers such as hedged fixed income information ratio of 0.5 and currency information ratio of 0.3, using a currency specialist with a currency information ratio of 1.0 will double the total information ratio.