Papers

Another look at the information ratio

Received (in revised form): 11th October, 2007

Ludwig B. Chincarini*

is an assistant professor of economics at Pomona College, as well as an adjunct professor of finance at Georgetown University, a financial consultant to money managers and hedge funds, and a member of the advisory board of IndexIQ. Prior to this, he was Director of Research at the index mutual fund company, Rydex Global Advisors, where he co-developed the new S&P 500 Equal-Weight index in collaboration with Standard & Poors. He also served a key role in launching the first equal-weight S&P 500 ETF. Before that he served as Director of Research for FOLIO*fn*. FOLIO*fn* was a revolutionary new brokerage firm that was one of the first brokerages to offer basket trading. Prior to building FOLIO*fn*, he worked at the Bank for International Settlements and at Schroders Asset Management. He graduated from the University of California at Berkeley with highest distinction, earning a BA in economics and holds a PhD in economics from MIT with a specialisation in macroeconomics and finance.

Daehwan Kim

is a senior portfolio manager at First Private Investment Management KAG. Prior to this, he was an assistant professor at the American University in Bulgaria. Before this, he worked for FOLIO*fn* as a financial economist, helping to create innovative financial products. He also worked for a business newspaper covering financial and other news, and had regular columns on financial economics in a business weekly. He earned his BA in economics from Seoul National University and his PhD in economics from Harvard University with a specialisation in econometrics and finance.

*Department of Economics, Pomona College, 425 N. College Avenue, #208, Claremont, California 91711, USA.
Tel: +1 909 621 8881; Fax: +1 202 687 4031; E-mail: chincarinil@hotmail.com

Abstract Chincarini and Kim (2006) argued that the information ratio can be interpreted as the square root of R^2. In this paper, we further develop this argument by, first, making a distinction between the conditional and the unconditional information ratio and, then by clarifying the relationship between R^2 and two versions of the information ratio. This paper also discusses the implications of our approach for interpreting the Fundamental Law of Active Management.
Journal of Asset Management (2007) **8,** 284–295. doi:10.1057/palgrave.jam.2250085

Keywords: *information ratio, fundamental law of active management, linear regression*

Introduction

The information ratio is a popular measure of risk-adjusted return performance for active portfolio managers. The numerator of this ratio is the active return or 'alpha', (α) that is, the constant term in the regression of the portfolio return on the benchmark return. The denominator is the active risk or 'omega' (ω), that is, the standard deviation of the error in the same regression. Through the fundamental law of active management, Grinold (1989) and Grinold and Kahn (1999) showed that the information ratio is approximately related to the breadth (the number of distinct signals by which portfolio managers forecast asset returns) and the information coefficient (the average quality of those signals).[1] They also noted that the information ratio is equivalent to the maximum Sharpe ratio under certain conditions.[2]

Owing to the popularity of the information ratio and the fundamental law among practitioners, a number of papers have been written recently on this topic. Relatively more attention has focused on the right-hand side of the fundamental law, that is the breadth and the information coefficient. For example, Clarke *et al.* (2002) extended the fundamental law to the case where additional constraints are added to the portfolio optimisation. Buckle (2004) extended the model of Grinold and Kahn so that the breadth can be measured without the knowledge of the number of signals.[3] Goodwin (1998), on the other hand, focused on the information ratio itself. He made an excellent exposition of various interpretations of the information ratio, and he especially clarifies the relationship between the information ratio and the Sharpe ratio, and between the information ratio and the *t*-statistic. Chincarini and Kim (2006) looked at the information ratio from an econometric perspective. They argued that the information ratio can be interpreted as the square root of R^2, the standard measure of goodness of fit in regression analysis.

In this paper, we take another look at the information ratio, and further develop the argument of Chincarini and Kim (2006).[4] We will make a distinction between the conditional information ratio and the unconditional information ratio. The conditional information ratio is a useful theoretical concept; but it is not something that can be estimated from real data and used to construct portfolios. The unconditional information ratio is what is important for portfolio construction, since it can be estimated from actual data. We will show that the squared unconditional information ratio equals the ratio of the regression sum of squares (RSS) to the error sum of squares (ESS). The ratio of RSS to ESS, in turn, can be approximated by the ratio of RSS to the total sum of squares (TSS).[5] This resulting ratio can be interpreted as a generalised R^2, a measure of the goodness of fit in a multivariate regression analysis.

Our interpretation of the information ratio has a number of practical advantages. First of all, our interpretation is fairly simple and intuitive to those familar with basic econometrics. It is consistent with the intuition that the portfolio manager's value added is directly related to the power of his or her forecasting ability, that is, the power of the underlying regression. The simplest measure of the power of a regression is the goodness of fit.

Second, and perhaps a more important advantage of our interpretation, is that it provides an alternative perspective on the fundamental law. The fundamental law breaks down the portfolio manager's value added into his or her breadth and skills. Our view of the fundamental law is that the portfolio manager's value added comes from the power of the regression analysis, and the power of the regression analysis can be broken down into the breadth of the regression and the quality of the forecasting variables.

This perspective leads us to formulate the breadth and the information coefficient in a new way. Once the nature of the information ratio is recognised, it is easy to see that the *breadth* and the *information coefficient* can be defined in a simple, measurable way. We argue that the way to measure breadth is to count the number of linearly independent forecasting variables. We justify our proposal by showing that the 'true' number of distinct signals cannot be determined *ex post*, and that signals affect the power of forecasting *only through forecasting variables.* We believe this perspective may be beneficial to understanding active management, since currently breadth, the number of distinct signals, is not measurable in practice.[6]

The remainder of the paper consists of four sections. The next section explains the conditional and the unconditional information ratio, and shows that the unconditional information ratio is approximately equal to the square root of the generalised R^2. The subsequent section illustrates the concepts with a simple example. The penultimate section discusses the implications of this for the fundamental law. The last section concludes.

The conditional IR, the unconditional IR, and the generalised R^2

Grinold and Kahn's framework can be best understood as a linear regression model. We first explain Grinold and Kahn's framework as a linear regression model, and spell out the assumptions that they made implicitly as well as explicitly. After that, we define the conditional information ratio and the unconditional information ratio and examine the relationship between the information ratio and R^2.

The linear regression model

Grinold and Kahn made three sets of assumptions to derive what is known as the fundamental law. First, Grinold and Kahn assumed that the return generating process is a linear function of certain exogenous variables. Secondly, they assumed that the capital asset pricing model is correct for the given benchmark. That is, the expected return of individual stocks is determined as the product of the beta and the benchmark return. Thirdly, Grinold and Kahn assumed that the linear regression model does not have any predictive power for the benchmark return. That is, the exogenous variables can help to predict individual stock returns, but not the benchmark return. We discuss each of these assumptions in turn.

Suppose that the investment universe is comprised of N stocks whose return is denoted by vector $\mathbf{r}$. We assume that the return generating process for $\mathbf{r}$ is a linear function of exogenous variables $\mathbf{x}$. Formally, we assume the following linear regression model for individual stock returns.

Assumption 1 (Return Generating Process). Stock return vector $\mathbf{r}$ is generated from the following linear regression model:

$$\mathbf{r} = \mathbf{a} + \mathbf{B}\mathbf{x} + \boldsymbol{\nu} \tag{1}$$

where

$$\begin{aligned} E(\boldsymbol{\nu}|\mathbf{x}) &= 0 \\ V(\boldsymbol{\nu}|\mathbf{x}) &= \mathbf{S} \end{aligned} \tag{2}$$

and

$$E(\mathbf{x}) = 0 \tag{3}$$

where $\mathbf{S}$ is a symmetric positive definite matrix.

In Equation (1), $\mathbf{x}$ represents the 'surprise' or the 'shock'; $\boldsymbol{\nu}$ is 'noise' that influences the stock returns randomly. Equation (2) implies that the mean of stock returns *conditional* on the value of $\mathbf{x}$ is $\mathbf{a} + \mathbf{B}\mathbf{x}$. It also implies that the variance of stock returns *conditional* on the value of $\mathbf{x}$ is $\mathbf{S}$. Given that $\mathbf{x}$ represents 'surprise', it is natural to assume that its expected value is zero, as indicated in Equation (3). Thus, $\mathbf{a}$ is the unconditional mean of stock returns.

The next assumption used by Grinold and Kahn has to do with the capital asset pricing model. The capital asset pricing model implies that, given the benchmark return r_B (which we assume to be the market portfolio with all the ideal properties), the returns of individual stocks can be written in the following way:

$$\mathbf{r} = \boldsymbol{\beta} r_B + \boldsymbol{\varepsilon} \tag{4}$$

where $\boldsymbol{\beta} r_B$ is the part of $\mathbf{r}$ correlated with the benchmark, and $\boldsymbol{\varepsilon}$ is the uncorrelated part. $\boldsymbol{\varepsilon}$ is often called the active return, as it is the component of return, not coming from the simple correlation with the benchmark, but possibly resulting from active management. Denoting the benchmark weights as $\mathbf{w}_B$ so that $r_B = \mathbf{w}_B'\mathbf{r}$,

$$\begin{aligned} \boldsymbol{\beta} &= \frac{V(\mathbf{r})\mathbf{w}_B}{\mathbf{w}_B' V(\mathbf{r})\mathbf{w}_B} \\ \boldsymbol{\varepsilon} &= [\mathbf{I} - \boldsymbol{\beta}\mathbf{w}_B'] r \end{aligned} \tag{5}$$

The capital asset pricing model requires that the expected value of the active return is zero, that is $E(\boldsymbol{\varepsilon}) = 0$. Given that the unconditional mean of stock returns is $\mathbf{a}$, we can write their second assumption more succinctly as

Assumption 2 (Capital Asset Pricing Model). The capital asset pricing model is assumed to be correct in the sense that $E(\mathbf{r}) = \boldsymbol{\beta} E(r_B)$.

This implies that the constant term in the return generating process, that is $\mathbf{a}$ in Equation (1), must satisfy the following constraint:

$$[\mathbf{I} - \boldsymbol{\beta}\mathbf{w}'_B]\mathbf{a} = 0 \tag{6}$$

That is, the capital asset pricing model imposes a restriction that the unconditional mean of stock returns (**a**) is orthogonal to a function of $\boldsymbol{\beta}$ and $\mathbf{w}_B$. Given this restriction, we can express the active return $\boldsymbol{\varepsilon}$ as following:

$$\boldsymbol{\varepsilon} = [\mathbf{I} - \boldsymbol{\beta}\mathbf{w}_B'](\mathbf{Bx} + \boldsymbol{\nu}) \tag{7}$$

The third assumption made by Grinold and Kahn was that the linear regression model does not have any predictive power for the benchmark return. Unlike the first two assumptions, this assumption is not of substantial value. That is, we could derive our main result without making this assumption. With this assumption, however, we can focus solely on active returns rather than total returns. This third assumption is summarised in the following statement.

Assumption 3 (No Prediction of Benchmark Return). The linear regression model does not have any predictive power for the benchmark return.

This implies that the 'slope' term in the return generating process, that is, **B** in Equation (1), must satisfy the following constraint:

$$\mathbf{w}_B'\mathbf{B} = 0 \tag{8}$$

This assumption simplifies the active return further. Now, we can express the active return $\boldsymbol{\varepsilon}$ as:

$$\boldsymbol{\varepsilon} = \mathbf{Bx} + \boldsymbol{\nu}^* \tag{9}$$

where

$$\begin{aligned} E(\boldsymbol{\nu}^*|\mathbf{x}) &= 0 \\ V(\boldsymbol{\nu}^*|\mathbf{x}) &= \mathbf{S}^* \end{aligned} \tag{10}$$

and $\boldsymbol{\nu}^*$ is $(\mathbf{I}-\boldsymbol{\beta}\mathbf{w}_B')\boldsymbol{\nu}$, while $\mathbf{S}^*$ is $(\mathbf{I}-\boldsymbol{\beta}\mathbf{w}_B')\mathbf{S}(\mathbf{I}-\mathbf{w}_B\boldsymbol{\beta}')$. The reader should notice a similarity between Equation (1) and Equation (9). Basically, assumptions 2 and 3 transformed the total return generating process into an active return generating process, while maintaining the basic structure of the linear regression model intact. The active returns are again a linear function of $\mathbf{x}$, and the new error term $\boldsymbol{\nu}^*$ is orthogonal to $\mathbf{x}$. The intuition is simple. Since the linear regression model for total returns does not say anything about the benchmark return, we can use the same linear regression model for the active return part only.

There is a peculiarity in the linear regression model for active returns. From Assumption 3 and the definition of $\mathbf{S}^*$, $\boldsymbol{\varepsilon}$ is orthogonal to $\mathbf{w}_B$. That is, elements of $\boldsymbol{\varepsilon}$ are not linearly dependent, and only $N-1$ elements of $\boldsymbol{\varepsilon}$ are linearly dependent. In practical terms, we cannot estimate Equation (9) as it is. We can only estimate $N-1$ elements of $\boldsymbol{\varepsilon}$ by a linear regression. Thus, we *partition* Equation (9) into two parts: the equation for the first element of $\boldsymbol{\varepsilon}$ (denoted by ε_1) and the equation for the remainder of $\boldsymbol{\varepsilon}$ (denoted by $\boldsymbol{\varepsilon}_2$). We can apply the linear regression model for $\boldsymbol{\varepsilon}_2$, and ε_1 will be determined automatically from $\boldsymbol{\varepsilon}_2$. For future reference, the linear regression model for $\boldsymbol{\varepsilon}_2$ is

$$\boldsymbol{\varepsilon}_2 = \mathbf{B}_2\mathbf{x} + \boldsymbol{\nu}_2^* \tag{11}$$

where

$$\begin{aligned} E(\boldsymbol{\nu}_2^*|\mathbf{x}) &= 0 \\ V(\boldsymbol{\nu}_2^*|\mathbf{x}) &= \mathbf{S}_{22}^* \end{aligned} \tag{12}$$

and $\mathbf{B}_2$ is created by eliminating the first row from $\mathbf{B}$. $\mathbf{S}_{22}^*$ is created by eliminating the first row and the first column from $\mathbf{S}^*$.

The conditional IR and the unconditional IR

A given portfolio return r_B can be decomposed into two parts.[7] That is,

$$r_P = \beta_P r_B + \varepsilon_P \tag{13}$$

where $\beta_P r_B$ is the part of the portfolio return correlated with the benchmark and ε_P is the active return of the portfolio. Let us denote the vector of portfolio weights as $\mathbf{w}_P$, such that $r_P = \mathbf{w}_P'\mathbf{r}$. Then $\beta_P = \mathbf{w}_P{}'\boldsymbol{\beta}$ and $\varepsilon_P = \mathbf{w}_P'\boldsymbol{\varepsilon}$.Thegoaloftheactivemanageris to find the maximum information ratio, which is obtained by finding the maximum ratio of the expected value of active return to the risk of the active return.

Before moving on, let us examine some characteristics of Equation (9). The randomness of ε_P arises from the exogenous variables $\mathbf{x}$ and the random error $\boldsymbol{\nu}^*$. While $\boldsymbol{\nu}^*$ is never observed, $\mathbf{x}$ can be observed. This means that we have two ways to calculate the expected value of ε_P. We could calculate the unconditional expected value of ε_P. Or we could calculate the expectation of ε_P conditional on the value of $\mathbf{x}$. However, the unconditional expectation of ε_P is not interesting to us as we know that it is always zero. Thus, it makes sense to calculate the conditional expectation of ε_P rather than the unconditional expectation of ε_P.

Let us denote the expected value and the risk of the active return conditional on the value of $\mathbf{x}$ as α_P and ω_P, respectively. Then

$$\begin{aligned} \alpha_P &= E(\varepsilon_P|\mathbf{x}) = \mathbf{w}'_P\mathbf{B}\mathbf{x} \\ \omega_P^2 &= V(\varepsilon_P|\mathbf{x}) = \mathbf{w}'_P\mathbf{S}^*\mathbf{w}_P \end{aligned} \tag{14}$$

We call the maximum ratio of α_P to ω_P *the conditional information ratio*[8]

$$\mathrm{IR}_c \equiv \max \frac{\alpha_P}{\omega_P} \tag{15}$$

It is the conditional information ratio as its value is conditional on some value of $\mathbf{x}$. When the value of $\mathbf{x}$ is not known, we cannot determine the conditional information ratio. In such a case, we can take the expectation of the conditional information ratio with respect to $\mathbf{x}$.[9] We call the expectation of the conditional information ratio *the unconditional information ratio*. Formally,

$$\mathrm{IR}_u^2 \equiv E_{\mathbf{x}}(\mathrm{IR}_c^2) \tag{16}$$

where the subscript to the function E indicates that the expectation is taken with respect to $\mathbf{x}$. Equation (16) is really an equation for the *squared* unconditional information ratio.[10] The unconditional information ratio is the central tendency of the conditional information ratio as the exogenous variables $\mathbf{x}$ take different values. There are some important distinctions between the conditional information ratio and the unconditional information ratio.

First, while the (unconditional) expected value of ε_P and α_P is zero, the expected value of the unconditional information ratio is not zero.[11]

Secondly, what appears on the left-hand side of the fundamental law is the unconditional information ratio. For this reason, the information ratio mentioned in the literature is mostly referring to the unconditional information ratio. However, when people discuss the '*ex post*' information ratio, sometimes they are in fact referring to an *estimate* of the conditional information ratio.

Thirdly, both the conditional information ratio and the unconditional information ratio are functions of *unknown* parameters. Thus, they may be estimated from the data, but their true value cannot be known. Any information ratio one might calculate from the data is in fact an estimate.

Fourthly, we do not want to equate the conditional/unconditional distinction with the *ex ante/ex post* distinction. The exact meaning of *ex ante* and *ex post* is rather unclear, and people use these terms rather vaguely. Some people use the *ex ante/ex post* distinction to refer to the conditional/unconditional distinction, while others use them to refer to the estimator/estimate distinction.[12]

For future reference, we present an expression for the conditional information ratio based on the 'partitioned' linear regression model of the active returns, that is, Equation (11). Let us denote the first element of the benchmark weight vector and the portfolio weight vector as w_{B1} and w_{P1} and the remainder of the benchmark weight vector and the portfolio weight vector as $\mathbf{w}_{B2}$ and $\mathbf{w}_{P2}$. Similarly, let us denote the first row of matrix $\mathbf{B}$ as $\mathbf{B}_1$ and the remainder as $\mathbf{B}_2$. For $\mathbf{S}^*$, we will partition the matrix into four parts, the upper-left element $\mathbf{S}_{11}^*$, the upper-right part $\mathbf{S}_{21}^{*\prime}$, the lower-left part $\mathbf{S}_{21}^*$, and the lower-right part $\mathbf{S}_{22}^*$ (In this particular case, $\mathbf{S}_{11}^*$ is a scalar). From the orthogonality between the benchmark weights and $\mathbf{B}$, we can express $\mathbf{B}_1\mathbf{x}$ as $\mathbf{k}'\mathbf{B}_2$ where $\mathbf{k} = -\mathbf{w}_{B2}/\mathbf{w}_{B1}$. From the orthogonality between the benchmark weights and $\mathbf{S}^*$, we can write $\mathbf{S}_{11}^*$ as $\mathbf{k}'\mathbf{S}_{22}^*\mathbf{k}$ and $\mathbf{S}_{21}^*$ as $\mathbf{S}_{22}^*\mathbf{k}$. Then

$$\begin{aligned}\alpha_P &= \tilde{\mathbf{w}}_{P2}'\mathbf{B}_2\mathbf{x} \\ \omega_P^2 &= \tilde{\mathbf{w}}_{P2}'\mathbf{S}_{22}^*\tilde{\mathbf{w}}_{P2}\end{aligned} \tag{17}$$

where $\tilde{\mathbf{w}}_{P2} = \mathrm{w}_{P1}\mathbf{k}' + \mathbf{w}_{P2}$ Then the conditional information ratio is

$$\mathrm{IR}_c = \max_{\tilde{\mathbf{w}}_{P2}} \frac{\tilde{\mathbf{w}}_{P2}'\mathbf{B}_2\mathbf{x}}{\tilde{\mathbf{w}}_{P2}'\mathbf{S}_{22}^*\tilde{\mathbf{w}}_{P2}} \tag{18}$$

In Equation (18) the control variable of the maximisation problem changed from $\mathbf{w}_P$ to $\tilde{\mathbf{w}}_{P2}$. The dimension of $\tilde{\mathbf{w}}_{P2}$ is smaller than $\mathbf{w}_P$. This is due to the fact that the active returns are orthogonal to the benchmark returns. We also do not impose the restriction that the sum of the portfolio weights should be one in the maximisation. If the sum of the weights is not one, we can simply rescale the weights, and this rescaling does not influence the ratio of α_P to ω_P. Thus, we will ignore this constraint in our analysis.[13]

The unconditional IR and the generalised R^2

We reformulate the maximisation problem of Equation (18) into a more familar-looking variance minimisation problem. We can minimise ω_P for some value of α_P. The solution to this problem will depend on the value of α_P. By varying the value of α_P, we obtain a set of $\tilde{\mathbf{w}}_{P2}$ that minimises the residual variance. From this set, we can find the maximum value of α_P/ω_P. The variance minimisation problem can be written as follows:

$$\begin{aligned}&\min_{\tilde{\mathbf{w}}_{P2}} \tilde{\mathbf{w}}_{P2}'\mathbf{S}_{22}^*\tilde{\mathbf{w}}_{P2} \\ &s.t. \quad \tilde{\mathbf{w}}_{P2}'\mathbf{B}_2\mathbf{x} = \mu\end{aligned} \tag{19}$$

where μ is a scalar, whose value we want to vary to generate the set of *optimal* weight vectors.

The solution $\tilde{\mathbf{w}}_{P2}^*$ and the corresponding α_P/ω_P are

$$\begin{aligned}\tilde{\mathbf{w}}_{P2}^* &= \frac{\mu\mathbf{S}_{22}{}^{*-1}\mathbf{B}_2\mathbf{x}}{\mathbf{x}'\mathbf{B}_2'\mathbf{S}_{22}^{*-1}\mathbf{B}_2\mathbf{x}} \quad \frac{\alpha_P}{\omega_P} \\ &= \sqrt{\mathbf{x}'\mathbf{B}_2'\mathbf{S}_{22}^{*-1}\mathbf{B}_2\mathbf{x}}\end{aligned} \tag{20}$$

The equation for the value α_P/ω_P does not depend on μ. That is, the ratio of α_P to ω_P is constant regardless of the target mean μ, as long as the minimum variance is reached. Thus we can find the conditional information ratio without further calculation:

$$\mathrm{IR}_c = \sqrt{\mathbf{x}'\mathbf{B}_2'\mathbf{S}_{22}^{*-1}\mathbf{B}_2\mathbf{x}} \tag{21}$$

The squared unconditional information ratio is obtained by integrating out $\mathbf{x}$:

$$\mathrm{IR}_u^2 = E\sqrt{\mathbf{x}'\mathbf{B}_2'\mathbf{S}_{22}^{*-1}\mathbf{B}_2\mathbf{x}} \tag{22}$$

The squared unconditional information ratio obtained above can be interpreted as the generalised R^2 of the linear regression model for the active returns $\boldsymbol{\varepsilon}_2$. In a regression with one dependent variable, R^2 is defined as the ratio of the RSS to the TSS, where RSS is the variation in the explanatory variables and their coefficients and TSS is the variation in the dependent variable. That is, given the regression model

$$y = bz + \eta \tag{23}$$

R^2 is defined as

$$R^2 = \frac{V(bz)}{V(y)} \tag{24}$$

We can generalise R^2 for the regression with many dependent variables in a straightforward way. We define TSS as the variance–covariance matrix of the dependent variables, and RSS as the variance–covariance matrix of the explanatory variables and their coefficients. Mathematically, we just replace scalars in Equations (23) and (24) with vectors and matrices. That is, given the regression model

$$\mathbf{y} = \mathbf{bz} + \boldsymbol{\eta} \tag{25}$$

the *generalised* R^2 is defined as[14]

$$R^2 = \mathrm{tr}\left(V(\mathbf{bz})^{\frac{1}{2}} V(\mathbf{y})^{-1} V(\mathbf{bz})^{\frac{1'}{2}} \right) \tag{26}$$

Equations (22) and (26) are very similar in structure. The similarity is not an accident. We can approximate the squared unconditional information ratio IR_u^2 as the generalised R^2 of the regression of the model in Equation (11). Starting from the generalised R^2 of Equation (11), we have

$$\begin{aligned} R^2 =& \mathrm{tr}[V(\mathbf{B}_2\mathbf{x})^{\frac{1}{2}} V(\boldsymbol{\varepsilon}_2)^{-1} V(\mathbf{B}_2\mathbf{x})^{\frac{1}{2}\prime}] \\ =& \mathrm{tr}[V(\boldsymbol{\varepsilon}_2)^{-1} V(\mathbf{B}_2\mathbf{x})] \\ \approx& \mathrm{tr}[V(\boldsymbol{\varepsilon}_2|\mathbf{x})^{-1} V(\mathbf{B}_2\mathbf{x})] \\ =& \mathrm{tr}[\mathbf{S}_{22}^{*-1}\mathbf{B}_2\mathrm{E}(\mathbf{xx}')\mathbf{B}'_2] \\ =& E(\mathbf{x}'_2\mathbf{S}_{22}^{*-1}\mathbf{B}_2\mathbf{x}) = \mathrm{IR}_u^2 \end{aligned} \tag{27}$$

Only one approximation was made in the above derivation. The approximation was to replace $V(\boldsymbol{\varepsilon}_2)$ with $V(\boldsymbol{\varepsilon}_2|\mathbf{x})$. In the regression framework, we are assuming that RSS is small relative to TSS. That is, we are assuming that the R^2 is not too large.[15] If this is true, then replacing TSS in the denominator with the ESS will be approximately correct. This approximation is not of our own invention. It is the same approximation used by Grinold and Kahn to derive the fundamental law.

One might not understand how the generalised R^2 and the information ratio can be related since the former is bounded by 0 and 1, while the information ratio can take on an unlimited possible values. This can be explained by the fact that the two are only approximately related. In particular, when the ratio of RSS to TSS is significantly different from 0, then this relationship no longer holds. However, when the assumption that RSS is small relative to TSS is not true, neither is the fundamental law. That is, our approximation error is of the same size as that of the fundamental law.

Thus, if you are assuming that the fundamental law is valid and using it, you can also use our formulation that the generalised R^2 is equal to the squared information ratio.

An example

To illustrate these concepts in a more practical form, we will go through a realistic example consisting of a small portfolio and show how the generalised R^2 of the regression of stock returns is very close to the unconditional information ratio squared. Suppose that we are creating an equity portfolio consisting of five large companies — General Electrics (GE), Microsoft (MSFT), Pfizer (PFE), Wall-Mart (WMT), and Exxon Mobil (XOM). The benchmark is the value-weighted index of these five stocks. We want to use the following three variables — the change in consumer sentiment index, the size premium, and the value premium — for forecasting stock returns. Let us say that we are forming a portfolio at the end of December 1999, and that we would like to use five years of data to estimate our model.[16]

In this situation, it is natural to estimate the equation relating the return of each of the five stocks to three forecasting variables. However, if one believes that the benchmark cannot be predicted (as stipulated in Assumption 3), then one would estimate an equation for the residual (ie active return) rather than for the return. To obtain the active return, we estimate the following equations:

$$\begin{aligned} r_{\mathrm{GE},t} =& \alpha_{\mathrm{GE}} + \beta_{\mathrm{GE}} r_{B,t} + \varepsilon^*_{\mathrm{GE},t} \\ r_{\mathrm{MSFT},t} =& \alpha_{\mathrm{MSFT}} + \beta_{\mathrm{MSFT}} r_{B,t} + \varepsilon^*_{\mathrm{MSFT},t} \\ r_{\mathrm{PFE},t} =& \alpha_{\mathrm{PFE}} + \beta_{\mathrm{PFE}} r_{B,\mathrm{t}} + \varepsilon^*_{\mathrm{PFE},t} \\ r_{\mathrm{WMT},t} =& \alpha_{\mathrm{WMT}} + \beta_{\mathrm{WMT}} r_{B,t} + \varepsilon^*_{\mathrm{WMT},t} \\ r_{\mathrm{XOM},t} =& \alpha_{\mathrm{XOM}} + \beta_{\mathrm{XOM}} r_{B,t} + \varepsilon^*_{\mathrm{XOM},t} \\ t =& \mathrm{Jan}\ 1995, \ldots, \mathrm{Dec}\ 1999 \end{aligned} \tag{28}$$

where $r_{B,t}$ is the benchmark return. For each stock, we define the active return $\hat{\varepsilon}$ as the sum of the estimated α and the estimated ε^*.

That is,

$$
\begin{aligned}
\hat{\varepsilon}_{\mathrm{GE},t} &= \hat{\alpha}_{\mathrm{GE}} + \hat{\varepsilon}^{*}_{\mathrm{GE},t} \\
\hat{\varepsilon}_{\mathrm{MSFT},t} &= \hat{\alpha}_{\mathrm{MSFT}} + \hat{\varepsilon}^{*}_{\mathrm{MSFT},t} \\
\hat{\varepsilon}_{\mathrm{PFE},t} &= \hat{\alpha}_{\mathrm{PFE}} + \hat{\varepsilon}^{*}_{\mathrm{PFE},t} \\
\hat{\varepsilon}_{\mathrm{WMT},t} &= \hat{\alpha}_{\mathrm{WMT}} + \hat{\varepsilon}^{*}_{\mathrm{WMT},t} \\
\hat{\varepsilon}_{\mathrm{XOM},t} &= \hat{\alpha}_{\mathrm{XOM}} + \hat{\varepsilon}^{*}_{\mathrm{XOM},t} \\
t &= \text{Jan 1995}, \ldots, \text{Dec 1999}
\end{aligned} \tag{29}
$$

Once the active returns are obtained, we can estimate the equation relating the active returns to the three forecasting variables. Recall, however, from the previous section that it is not necessary to estimate the equation for every stock. One equation is redundant as active returns are not linearly independent. Thus, we drop the equation for GE, and estimate for the remaining four stocks only. That is,

$$
\begin{aligned}
{}_{\mathrm{MSFT},t} &= b_{\mathrm{MSFT},0} + b_{\mathrm{MSFT},1}x_{1,t} \\
&\quad + b_{\mathrm{MSFT},2}x_{2,t} + b_{\mathrm{MSFT},3}x_{3,t} \\
&\quad + v_{\mathrm{MSFT},t} \\
{}_{\mathrm{PFE},t} &= b_{\mathrm{PFE},0} + b_{\mathrm{PFE},1}x_{1,t} \\
&\quad + b_{\mathrm{PFE},2}x_{2,t} + b_{\mathrm{PFE},3}x_{3,t} + v_{\mathrm{PFE},t} \\
{}_{\mathrm{WMT},t} &= b_{\mathrm{WMT},0} + b_{\mathrm{WMT},1}x_{1,t} + b_{\mathrm{WMT},2}x_{2,t} \\
&\quad + b_{\mathrm{WMT},3}x_{3,t} + v_{\mathrm{WMT},t} \\
{}_{\mathrm{XOM},t} &= b_{\mathrm{XOM},0} + b_{\mathrm{XOM},1}x_{1,t} + b_{\mathrm{XOM},2}x_{2,t} \\
&\quad + b_{\mathrm{XOM},3}x_{3,t} + v_{\mathrm{XOM},t} \\
t &= \text{Jan 1995}, \ldots, \text{Dec 1999}
\end{aligned} \tag{30}
$$

where $x_{1,t}, x_{2,t}, x_{3,t}$ represent the change in the consumer sentiment index, the size premium, and the value premium.[17] Table 1 presents the estimates of the coefficients and Table 2 presents the variance–covariance matrix estimates of the errors.

If we knew the values of $x_{1,t}, x_{2,t}, x_{3,t}$ for January 2000, we could use the above estimates and obtain the distribution of the active returns for January 2000. As in the previous section, let us denote the expected value and the standard deviation of the active portfolio return as α_P and ω_P By solving the minimisation problem of (19), we obtain the efficient frontier shown in Figure 1.

The exact efficient frontier is obtained only when we know the exact values of $x_{1,t}, x_{2,t}, x_{3,t}$. In practice, this is not possible, since the exact values of the factors are not known at the time of portfolio formation which in our case is the end of December 1999. For the purposes of illustration, we used the actual January 2000 values of $x_{1,t}, x_{2,t}, x_{3,t}$ to draw this frontier.

The slope of the figure is the conditional information ratio. It is conditional on the value of $x_{1,t}, x_{2,t}, x_{3,t}$. Using Equation (21) and the January 2000 values of $x_{1,t}, x_{2,t}, x_{3,t}$ (which, together with 1, makes $\mathbf{x}$ below)[18],

$$
\mathrm{IR}_c = \sqrt{\mathbf{x}'\hat{\mathbf{B}}'_2\hat{\mathbf{S}}_{22}^{*-1}\hat{\mathbf{B}}_2\mathbf{x}} = 0.8129 \tag{31}
$$

where we obtain the values of $\hat{\mathbf{B}}_2$ from Table 1 and the values for $\hat{\mathbf{S}}^{*}_{22}$ from Table 2.

When we do not know the values $x_{1,t}, x_{2,t}, x_{3,t}$ for January 2000, we need to calculate the unconditional information ratio. Using the third line of Equation (27),

$$
\begin{aligned}
\mathrm{IR}_u^2 &= \mathrm{tr}\left[\hat{\mathbf{S}}_{22}^{*-1}\hat{\mathbf{B}}_2\hat{V}(\mathbf{x})\hat{\mathbf{B}}'_2\right] \\
&= 0.2242
\end{aligned} \tag{32}
$$

This is the exact value of the unconditional information ratio. We can see that this value

Table 1 Coefficient estimates of the linear regression model

Constant	Consumer	Sentiment	Size	Value
MSFT	−0.3737	44.9487	6.96	−22.0888
PFE	0.3307	13.4944	−54.5847	6.1705
WMT	0.6035	−19.232	−2.1064	−14.0777
XOM	1.425	−7.7477	28.7622	24.5179

Note: Based on monthly data from January 1995 to December 1999. The dependent variables are in percentage form. That is, 10 indicates a 10 per cent return. The explanatory variables are not in per cent. Thus, a 1 per cent change in the consumer sentiment index leads to a 0.44 per cent change in the return of MSFT according to the model.

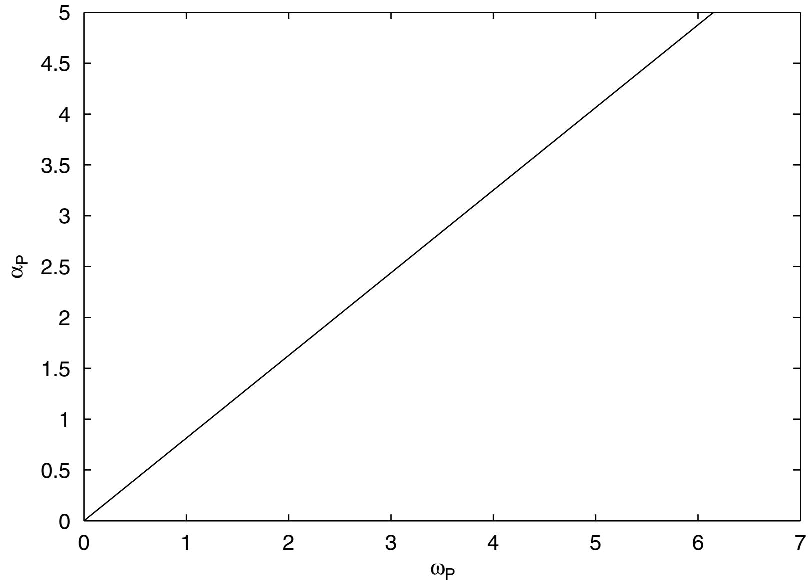

Figure 1 The efficient frontier in terms of α_P and ω_P

is not very different from the generalised R^2 of the regression, which is

$$R^2 = \mathrm{tr}[\hat{V}(\boldsymbol{\varepsilon}_2)\hat{\mathbf{B}}_2\hat{V}(\mathbf{x})\hat{\mathbf{B}}_2'] = 0.2042 \tag{33}$$

That is, in this particular example, the approximation error is about 0.02, and R^2 is a reasonable substitute for the squared unconditional information ratio.

Reinterpreting the fundamental law

Once we adopt the idea that the information ratio is approximately equal to R^2, the fundamental law can be interpreted in a very simple way. The breadth is the number of linearly independent explanatory variables in the linear regression model of active returns, that is Equation (11), and the information coefficient is the average contribution of these explanatory variables in increasing R^2.[19]

There are a group of people who might disagree with our view of the fundamental law. We will collectively refer to this group as our sceptics.[20] The sceptics might disagree with our view in two ways, first through the measurement of breadth, and second through the measurement of the information coefficient. We will discuss each of these objections in turn.

Table 2 Variance–covariance matrix estimates of the linear regression model

	MSFT	PFE	WMT	XOM
MSFT	34.6939			
PFE	−5.5209	37.7748		
WMT	−21.6087	−13.0777	47.7013	
XOM	−2.4188	−0.7983	−6.0301	17.6963

Note: Based on monthly data from January 1995 to December 1999. The dependent variables are in percentage form. That is, 10 indicates a 10 per cent return. The explanatory variables are not in per cent. Thus, in this table, MSFT has a residual variance of 0.3469 per cent per month.

Our sceptics argue that breadth should count the number of 'signals', not the number of explanatory variables. Consider, for example, a variable representing the consensus forecast of earnings per share. If this is the only variable used in the model, we would argue that the breadth is one. Our sceptics would claim that the number of 'signals' can be as high as a few thousand if the consensus forecast is based on thousands of analysts.

We reject our sceptics' view on two grounds. Firstly, the number of signals cannot be determined uniquely. Once the set of variable $\mathbf{x}$ is determined, there is no way to recover the number of signals from the distribution of $\mathbf{x}$. Neither the mean $E(\mathbf{x})$ nor the variance–covariance matrix $V(\mathbf{x})$ can tell us the number of signals. It is easy to show mathematically that given a vector of random variable $\mathbf{x}$, there are an unlimited number of orthonormal vectors $\mathbf{y}$ that satisfy $\mathbf{x} = \mathbf{G}\mathbf{y}$, and the dimension of $\mathbf{y}$ (which is the number of signals in Grinold and Kahn's framework) can be essentially anything.

Secondly, and more important, even if we know the exact number of signals generating the explanatory variable $\mathbf{x}$, it does not really matter. Whether the consensus forecast is made out of 1,000 analysts' forecasts or out of 100 analysts' forecasts is irrelevant. All that matters is the distribution of $\mathbf{x}$ (ie the mean and the variance in the normal distribution model). Given the distribution of $\mathbf{x}$, other information about how $\mathbf{x}$ was created does not influence either the information ratio or R^2. [21]

The second way our sceptics might have issues with our proposal is through the information coefficient. Our sceptics often treat the information coefficient as the average of the correlation between *a* stock return and *all* explanatory variables. We believe this is an inappropriate, if not incorrect, interpretation of the information coefficient. If one looks at the formula of Grinold and Kahn carefully, it is clear that the information ratio is the average of the correlation between *all* stock returns and *an* explanatory variable. The information coefficient shows the amount of information in each explanatory variable, not in each stock return.

Given an $N-1$ dimensional vector of active returns, $\boldsymbol{\varepsilon}_2$, and an M dimensional vector of explanatory variables, $\mathbf{x}$, one can calculate an $(N-1) \times M$ number of correlations. Grinold and Kahn first take the average across active returns for *each* explanatory variable, and obtain M average correlations.[22] Each of these M average correlations can be called an information coefficient, reflecting the amount of information in each explanatory variable. Only after this step, do Grinold and Kahn average the M information coefficients.[23] Our sceptics' view amounts to changing the order of the averaging, first they average across M explanatory variables, and then they average across $N-1$ stocks.

If one takes our view that the information coefficient reflects the correlation between an explanatory variable and all stocks, then one can see the intuition behind our claim that the information coefficient is the average contribution of explanatory variables in increasing R^2. The name R^2 comes from the fact that it is the squared correlation coefficient, which is often denoted by the Greek letter ρ. So it is natural to relate the information coefficient, which is essentially the correlation coefficient, with R^2.[24]

The fundamental law developed by Grinold and Kahn states that the information ratio equals the square root of breadth multiplied by the information coefficient. In equation form, it is written as:

$$\mathrm{IR} = \mathrm{IC}\sqrt{\mathrm{Breadth}} \tag{34}$$

Owing to the unobservable nature of signals, both breadth and the information coefficient remain ambiguous items to portfolio managers. We specify a regression framework in which actual explanatory variables are used in a regression model rather than signals. This allows us to make definitive statements about the fundamental law. First, the generalised R^2 of the regression of factors on stock returns is approximately equal to the unconditional information ratio. Secondly, the breadth of the fundamental law is equal to the number of explanatory variables used in the regression analysis, and thirdly, the information coefficient is simply the average contribution of the explanatory variables in determining R^2. In the example we provided in the previous section, the breadth is 3 and the information coefficient (IC) is 0.2609 $\left(\sqrt{0.2042/3}\right)$.

For portfolio managers who use linear regression models to form portfolios, this formulation should clarify the link between their models and the expected information ratios. For portfolio managers who do not use linear regression models, but use other techniques, this formulation might be useful in that it offers them another equivalent approach using regression analysis. For example, suppose a portfolio manager creates a portfolio based upon stocks whose analyst ratings increased over the last month. To understand the information ratio of such a portfolio, an equivalent approach would be to run a linear regression of stock returns against a dummy variable indicating whether the average analyst rating went up or down for that stock (ie a value of 1 for stocks whose average analyst rating went up and a value of 0 for stocks whose average analyst rating went down). And in this case, breadth and the information coefficient would be observable quantities, rather than an unknown, ambiguous concept.[25]

Conclusion

The information ratio is an important concept for active portfolio managers. Grinold and Kahn popularised a concept known as the fundamental law of active management which decomposes the information ratio into two components. One component is commonly known as *breadth* and the other component is commonly known as *the information coefficient*. These terms, especially breadth, however are not truly observable quantities. Thus, portfolio managers and analysts are confused with how to accurately compute breadth. We offer a new approach to the decomposition by linking a simple econometric model of security returns with the information ratio. We show that if security returns are indeed driven by portfolio factor models, the fundamental law can be interpreted in an observable and practical way. We do this by showing that the R^2 of the regression of security returns on factors is approximately equal to the squared information ratio.[26] That is, $IR^2 \approx R^2$. From this perspective, breadth and the information coefficient take on a very specific meaning. Breadth will always equal the number of factors that the portfolio manager or analyst uses to predict stock returns and the information coefficient is the average contribution of each factor to the forecasting regression. For example, if a portfolio manager believes that k factors predict stock returns and runs a linear regression of stock returns against these factor realisations over time, the R^2 of this regression is the best guess at the unconditional information ratio of the model and the breadth is k, while the information coefficient will equal $\sqrt{R^2/k}$.

Thus, practically portfolio managers can only improve the information ratio if they can improve the R^2 of the regressions, which can be done by finding more relevant or better predicting factors or by increasing the number of factors, without decreasing the average contribution of the factors. Beyond that, there is no special recipe to increase the information ratio.

Notes

1. The exposition of the fundamental law has a somewhat different 'flavour' in their 1989 article and in their 1999 book. We take the exposition of the 1999 book as our point of departure.
2. William Sharpe himself commented on this as well in his Sharpe (1994) article.
3. Both Clarke *et al.* (2002) and Buckle (2004) defined the information coefficient as the correlation between an asset return and the *aggregated* predictor of that return. That is, one information coefficient is defined for one asset. This is significantly different from how Grinold and Kahn defined the information coefficient in their 1999 book. In the book, Grinold and Kahn defined the information ratio as the correlation between an asset return and *one* forecasting variable. That is, in the model of Grinold and Kahn, one information coefficient is defined for one forecasting variable.
4. In Chincarini and Kim (2006), we did not provide a rigorous proof of this argument, partly due to the nature of the publication. The current paper is our first attempt to present a rigorous proof of the argument.
5. When we refer to two ratios being equal or approximately equal, we are referring to ratios in the matrix mathematical equivalent form.
6. While there was an attempt to make it measurable (eg Buckle, 2004), the resulting formula was rather complicated and not very intuitive.
7. This decomposition is similar to equation (4).
8. It is also the maximum information ratio, but we will not include this in the name from this point forward.
9. In mathematical jargon, we can 'integrate out' $\mathbf{x}$ and produce a formula that is independent of the value of $\mathbf{x}$.

10. This is of course different from defining the unconditional information ratio without taking a square. That is, $\sqrt{E_{\mathbf{x}}(\mathrm{IR}_c^2)}$ is different from $E_{\mathbf{x}}(\mathrm{IR}_c)$. Nonetheless, our main ideas are not affected by this technicality.
11. In general, the expected value of a function of X does not have to be zero even though the expected value of X is zero.
12. The *ex ante/ex post* distinction is used in at least two occasions. Some people use the *ex ante/ex post* terminology to distinguish 'an estimator' from 'an estimate'. An estimator is a formula whose value has not been determined yet. Once we have a real data set and calculate a specific value of an estimator, the specific value of the estimator is called an estimate. The IR is a formula, whose value needs to be computed from various parameters. Until specific values of the parameters are determined, a specific value of IR cannot be determined. At this stage, the IR is an estimator. Only after we obtain specific values for the parameters from a data set, can we determine a specific value for the IR. Once a specific value of the IR is determined, we may call this value an estimate. Some people use the term '*ex ante* IR' to refer to IR as an estimator, and '*ex post* IR' to refer to IR as an estimate. This distinction is independent from the conditional/unconditional distinction. If we take this particular interpretation of *ex ante* and *ex post*, we can think of '*ex ante* conditional IR' (conditional IR as an estimator), '*ex post* conditional IR' (estimate of conditional IR), '*ex ante* unconditional IR', (unconditional IR as an estimator) and '*ex post* unconditional IR' (an estimate of unconditional IR). Others use the *ex ante/ex post* terminology in a chronological sense. That is, if the real portfolio is not created yet, all the quantities calculated are *ex ante*. If the real portfolio has been already created, all the quantities calculated are *ex post*. This distinction is not identical to the conditional/unconditional distinction. What matters in the conditional/unconditional distinction is whether the latest values of the explanatory variables are known or not. Whether this is the case is independent of whether the real portfolio has already been created.
13. It is easy to show that the ratio of α_P to ω_P is not influenced when the portfolio weight vector is multiplied by a scalar.
14. If $\mathbf{y}$ is a scalar, equation (26) returns the standard R^2.
15. If we let $\mathrm{RSS/TSS} = x$, then the approximation error is $x^2/1-x$. For example, if x is about 1 per cent, then the approximation error is about 0.01 per cent. If x is about 10 per cent, then the approximation error is about 1 per cent, which is still reasonable.
16. In our particular example, we use data from January 1995 to December 1999.
17. For the size premium and the value premium, we use the variables created and distributed by Eugene Fama and Kenneth French.
18. In econometrics, the first column of the matrix is usually equal to one to account for the constant term in the regression.
19. The proof of this concept has already been presented in Chincarini and Kim (2006), thus we will focus on two aspects that were not fully explored by their work.
20. Grinold and Kahn (1999) might disagree with our view that breadth is simply the number of explanatory variables in the regression, rather than the number of signals. Clarke *et al*. (2002) and Buckle (2004) might disagree with our view that the information coefficient is the average of the correlation between all stock returns and an explanatory variable.
21. All that is important for portfolio construction in $\mathbf{x}$ is included in the distribution of $\mathbf{x}$. If $\mathbf{x}$ and $\mathbf{z}$ have the same distribution, it does not matter whether $\mathbf{x}$ and $\mathbf{z}$ are different variables. Both $\mathbf{x}$ and $\mathbf{z}$ will have an identical effect on the portfolio construction process. All computations will be identical. Thus, any other information we know about $\mathbf{x}$ other than the distribution is unnecessary as far as portfolio construction is concerned.
22. Grinold and Kahn calculated the correlations after orthonormalising $\boldsymbol{\varepsilon}_2$ and $\mathbf{x}$, but this orthonormalisation does not affect our argument, so we ignore it.
23. Grinold and Kahn assumed that the M information coefficients are identical. This is really not an assumption. What Grinold and Kahn really did was take the average of M information coefficients.
24. In general, the correlation among explanatory variables also matters in the calculation of R^2 and possibly IC. In the framework of Grinold and Kahn, all explanatory variables are independent from one another due to their orthogonalisation procedure. So the correlation among explanatory variables is zero. In our framework, this is not the case. However, the correlation among independent variables still do not influence average IC. This is so because the correlation among explanatory variables will cancel out one another through the averaging process.
25. We note that the fundamental law, both the new and old interpretation, is valid only if the portfolio manager creates an optimal portfolio. That is, improving breadth and the information coefficient may not result in higher performance if the portfolio is not optimal.
26. Our formulation is only an approximate relationship, as was the original Grinold and Kahn derivation; however, we added no new assumptions beyond their original assumptions.

References

Buckle, D. (2004) 'How to Calculate Breadth: An Evolution of the Fundamental Law of Active Portfolio Management', *Journal of Asset Management*, 4, 393–405.

Chincarini, L. B. and Kim, D. (2006) *Quantitative Equity Portfolio Management: An Active Approach to Portfolio Construction and Management*, McGraw-Hill: New York.

Clarke, R., Silva, H. de and Thorley, S. (2002) 'Portfolio Constraints and the Fundamental Law of Active Management', *Financial Analysts Journal*, 58, 48–66.

Goodwin, T. (1998) 'The Information Ratio', *Financial Analysts Journal*, 54, 34–43.

Grinold, R. C. (1989) 'The Fundamental Law of Active Management', *Journal of Portfolio Management*, 15, 30–37.

Grinold, R. C. and Kahn, R. N. (1999) *Active Portfolio Management: A Quantitative Approach to Providing Superior Returns and Controlling Risk*, McGraw-Hill: New York.

Sharpe, W. F. (1994) 'The Sharpe Ratio', *Journal of Portfolio Management*, 21, 49–59.

