On testing the equality of the multiple Sharpe Ratios, with application on the evaluation of iShares

Pui-lam Leung
Department of Statistics
The Chinese University of Hong Kong

and

Wing-Keung Wong
Risk Management Institute and Department of Economics
National University of Singapore

March 7, 2007

Acknowledgments

The authors would like to thank Professor Michael McAleer for his valuable comments. We also thank Professors Dominic Gasbarro and J. Kenton Zumwalt for providing us with the iShares data and for their helpful comments. The second author would like to thank Professors Robert B. Miller and Howard E. Thompson for their continuous guidance and encouragement. The research is partially supported by the grants from the Chinese University of Hong Kong and National University of Singapore.
On testing the equality of the multiple Sharpe Ratios, with application on the evaluation of iShares

Abstract:

Extending the work of Jobson and Korkie (1981), Lo (2002) and Memmel (2003), this paper applies the technique of the repeated measures design to develop the Multiple Sharpe ratio test statistic to test the hypothesis of the equality of the multiple Sharpe ratios. We also work out the asymptotic distribution of the statistic and its properties. To demonstrate the superiority of our proposed statistic over the traditional pair-wise Sharpe ratio test, we illustrate our approach by testing the equality of Sharpe ratios for the eighteen iShares. Whereas the pair-wise Sharpe ratio test show that the performance of all the 18 iShares are indistinguishable, our test results reject the equality of the Sharpe ratios in each year as well as in the entire sample; implying that the 18 iShares perform differently in each year as well as in the entire sample, with some outperforming others in the market. The test in our paper provides investors with a tool to evaluate their portfolio performances and enables them to make wiser decisions in their investments.

Keywords: MANOVA, Asymptotic distribution, $\delta$-method, iShares
1 Introduction

The basics of the mean-variance portfolio optimisation procedure have been well understood since the seminal work of Markowitz in the 1950’s. The Sharpe ratio (Sharpe 1966, 1994), the ratio of the excess expected return of an investment to its return volatility or standard deviation is one of the most commonly used statistics in the mean-variance framework. Originally motivated by the mean-variance analysis and the Sharpe-Linter Capital Asset Pricing Model, the Sharpe ratio is now used in many different contexts in Finance and Economics, from the evaluation of portfolio performance to tests of market efficiency for risk management (Levy 1972, Cumby and Glen 1990, Jorion 1991, Grinblatt and Titman 1994, Agarwal and Naik 2004). For example, Hodges, et al (1997) apply the Sharpe ratio to investigate the investment horizon of different portfolio types, namely small stocks, larger stocks and bonds. Results show that the Sharpe ratio first increases and then decreases in each portfolio type as the investment horizon lengthens and that bonds outperform stocks. Cochrane and Saa-Requejo (2000) impose the restrictions that investors would buy assets with high Sharpe ratio to pioneer option pricing by calculating the price bounds in one-period, multiperiod and continuous contexts. Lien (2002) finds portfolios with sufficiently large Sharpe ratios will have opposite rankings when tested by the Sortino ratio and the Upside Potential ratio. In addition, Leggio and Lien (2003) apply the Sharpe ratio as well as the Sortino ratio and the Upside Potential ratio to the dollar-cost averaging investment strategy and find that the relative ranking of dollar-cost averaging remains inferior to alternative investment strategies.

Although the Sharpe ratio has widespread use and myriad interpretations, little attention has been paid to its statistical properties. Because expected returns and volatilities can only be estimated, errors inevitably arise in the estimation of the Sharpe ratio. Jobson
and Korkie (1981) are the first to study the asymptotic distribution of empirical Sharpe Ratios and develop a test for the equality of two Sharpe Ratios, which is simplified by Memmel (2003). On the other hand, Lo (2002) derives a statistical distribution of the Sharpe ratio by using the standard econometric methods with several different sets of assumptions on the statistical behavior of the return series. With this statistical distribution, he shows that confidence intervals, standard errors, and hypothesis tests can be computed for the estimated Sharpe ratios in much the same way as regression coefficients such as portfolio alphas and betas are computed.

The Sharpe ratio test statistics developed by Jobson and Korkie (1981), Lo (2002) and Memmel (2003) are important as they provide a formal statistical comparison between the Sharpe ratios of two portfolios. However, it is common in Finance and Economics not only to make pairwise comparison, but also multiple comparison among a large number of stocks or portfolios but so far such test is not available. Thus, it is not surprise that many articles do not report their test results when they compare a large number of portfolios. For example, Gallagher and Jarnecic (2002) apply the Sharpe Ratio to analyse bond funds performance and conclude that active bond funds do not outperform the market benchmark. Edwards and Samant (2003) use the Sharpe ratio to evaluate risk-adjusted performance of socially responsible mutual funds during the period 1991 to 2000. So far, all the above studies do not report their testing results.

The unavailability of the multiple Sharpe ratios comparison calls forth alternative approaches. For example, Ackermann, et al (1999) compute the mean Sharpe Ratio and the median Sharpe ratio of several funds and test whether these estimates are significantly different from zero. Maller and Turkington (2002) compute the maximum Sharpe ratio from the assets and study the properties of such measure. These approaches enable investors to obtain some information about their portfolios but they cannot make multiple
comparison among all the Sharpe ratios nor take into consideration the joint effect of all the portfolios. This paper addresses this issue by developing the multivariate test statistic to test the hypothesis of the equality of the multiple Sharpe ratios. We also work out the asymptotic distribution of the statistic and its properties.

To verify the superiority of our test over the traditional pair-wise Sharpe ratio test, we apply both tests to 17 iShares and Standard and Poor’s Depository Receipts (SPY), which is treated as the eighteenth iShare in this paper. In March 1996, seventeen exchange traded funds (ETFs), now known as iShares, began trading on the American Stock Exchange and are used to track the Morgan Stanley Capital International (MSCI) foreign stock market indices. Comparison of the performance of different indices or funds are very important in finance literature. For example, Sharpe (1966) develops the Sharpe Ratio to compare 34 mutual funds with the Dow Jones Industrial Average (the Dow). He reports that between 1954-1963, 11 funds outperformed the Dow while 23 were outperformed by the Dow. Many studies use stock market indices that are not actually tradable or marketable. For example, daily returns are used by Peiro (1999) from nine stock indices while Jondeau and Rockinger (2003) utilize 20 international stock market indices. Prakash, et al (2003) employ both weekly and monthly returns from 17 international stock market indices. The indices used in these articles are not tradable. As iShares are tradable and mimic international indices, the introduction of iShares allows an investigation of realizable return of stock markets from different countries.

Applying the traditional pair-wise Sharpe ratio test, we fail to reject the equality of the Sharpe ratios for any pair of the iShares, This implies that the performances of all the 18 iShares is indistinguishable. However, as the pair-wise test does not take into consideration the joint effect of all the iShares, evaluation is non-conclusive. Then we apply the multivariate test developed in this paper to the 18 iShares. Contrary to the
pair-wise test, the Sharpe ratios of some iShares are different from the others for each year as well as for the entire sample. This means that some iShares outperform other iShares in the market. We note that the conclusion drawn from our findings is consistent with the findings in Gasbarro, et al (2006) who apply a stochastic dominance procedure to identify the dominance among iShares. The test developed in our paper provide more information in the evaluation of the portfolios’ performance and enables investors to make wiser decisions in their investment.

In Section 2, we describe the asymptotic distribution of the multiple Sharpe ratios and in Section 3 we lay out the multivariate test statistic for testing the hypothesis of the equality of multiple Sharpe ratios. In Section 4, we illustrate our approach by examining the equality of Sharpe ratios of the 18 iShares and in Section 5 we conclude our findings.

2 Asymptotic distribution of the Sharpe Ratio

Recall that the Sharpe ratio is defined as the ratio of the excess expected return to the standard deviation of return where the excess expected return is usually computed relative to the risk-free rate, \( r_f \). Consider \( k \) (\( k \geq 2 \)) portfolios of expected excess returns \( X_{1t}, \ldots, X_{kt} \) at time \( t \) with means \( \mu_1, \ldots, \mu_k \), variance \( \sigma^2_1, \ldots, \sigma^2_k \) and covariance, \( \sigma_{ij} \), of portfolio \( i \) and \( j \). For simplicity, in this paper we assume the excess returns to be serially independent and identically distributed (iid) as normal distribution and not subject to change through time.

We note that for many economic or financial data, \( \{X_{it}\} \) may not be iid. For example, it is well known that stock returns could be heteroskedastic in their variances. To circumvent this problem, it is common to include a Generalized Autoregressive Conditional
Heteroscedasticity (GARCH) innovation to allow for both autoregressive and moving average components in the heteroskedastic variance of the return to display a high degree of persistence.\(^1\) However, in this situation, one could easily transform them to be iid. For example, if \(\{X_{it}\}\) follow a GARCH model such that

\[
X_{it} = \mu_i + \varepsilon_{it} \tag{1}
\]

for \(i = 1, \ldots, k\) and \(t = 1, \ldots, T\); where \(\varepsilon_{it}\) follows a Generalized Autoregressive Conditional Heteroscedasticity (GARCH) \((p, q)\) model such that

\[
\begin{align*}
\varepsilon_{it} &= z_{it} \sqrt{h_{it}} \\
z_{it} &\sim iid \ N(0, 1) \\
h_{it} &= w_i + \sum_{j=1}^{p} a_{i,j} \varepsilon_{i,t-j}^2 + \sum_{k=1}^{q} b_{i,k} h_{i,t-k}.
\end{align*}
\]

Then, one could easily estimate \(\{h_{it}\}\) by \(\{\hat{h}_{it}\}\), transfer \(\{X_{it}\}\) into \(\{X^*_{it}\}\) by setting

\[
X^*_{it} = \frac{X_{it}}{\hat{h}_{it}}
\]

and thus \(\{X^*_{it}\}\) are iid for each \(i\). Thus, for simplicity and tractability, we assume \(\{X_{it}\}\) to be iid for each \(i\) in this paper.

Let \(\theta = (\mu_1, \ldots, \mu_k, \sigma^2_1, \ldots, \sigma^2_k)’\) be a \(2k \times 1\) vector of unknown parameters, the vector

\(^1\)see, for example, Bollerslev (1986), Bollerslev, et al. (1992, 1994), Diebold and Lopez (1995), Li, et al. (2002) and McAleer (2005) for more discussion.
of the Sharpe ratios of the $k$ portfolios is then an $k \times 1$ vector such that

$$u(\theta) = (r_1, \ldots, r_k)' = \left( \frac{\mu_1}{\sigma_1}, \ldots, \frac{\mu_k}{\sigma_k} \right)'.$$  

(2)

Going beyond Jobson and Korkie (1981), Lo (2002) and Memmel (2003), our test statistic test the hypothesis, $H_0$, of the equality of multiple Sharpe ratios such that:

$$H_0 : r_1 = \cdots = r_k.$$  

(3)

Let $\hat{\theta} = (\bar{x}_1, \ldots, \bar{x}_k, s_1^2, \ldots, s_k^2)'$ be the vector of the usual sample means and sample variances calculated from $n$ observations. The parameter $u(\theta)$ in (2) can then be estimated by

$$u(\hat{\theta}) = \left( \frac{\bar{x}_1}{s_1}, \ldots, \frac{\bar{x}_k}{s_k} \right)'.$$  

(4)

According to the standard asymptotic distribution theory, we have

$$\sqrt{n} [\hat{\theta} - \theta] \xrightarrow{D} N(0, \Sigma)$$  

(5)
where $\Sigma = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix}$ is an $2k \times 2k$ matrix such that

$$
\Sigma_1 = \begin{pmatrix} \sigma_1^2 & \cdots & \sigma_{1k} \\ \vdots & \ddots & \vdots \\ \sigma_{k1} & \cdots & \sigma_k^2 \end{pmatrix},
$$

$$
\Sigma_2 = \begin{pmatrix} 2\sigma_1^4 & \cdots & 2\sigma_{1k}^2 \\ \vdots & \ddots & \vdots \\ 2\sigma_{k1}^2 & \cdots & 2\sigma_k^4 \end{pmatrix}
$$

and 0 is an $k \times k$ matrix of zeros.

Applying the multivariate version of the $\delta$-method, (see for example, p.492 of Bishop, Fienberg and Holland 1975), we have

$$
\sqrt{n}[u(\hat{\theta}) - u(\theta)] \xrightarrow{D} N \left[ 0, \left( \frac{\partial u}{\partial \theta} \right) \Sigma \left( \frac{\partial u}{\partial \theta} \right)' \right]
$$

(6)

where $(\partial u/\partial \theta) = [D_1 : D_2]$ is an $k \times 2k$ matrix with $D_1$ and $D_2$ to be $k \times k$ diagonal matrices whose diagonal elements are $(1/\sigma_1, \ldots, 1/\sigma_k)$; and $[-\mu_1/(2\sigma_1^3), \ldots, -\mu_k/(2\sigma_k^3)]$ respectively. It is straightforward to show that

$$
\left( \frac{\partial u}{\partial \theta} \right) \Sigma \left( \frac{\partial u}{\partial \theta} \right)' = D_1 \Sigma_1 D_1' + D_2 \Sigma_2 D_2'
$$

(7)
where

\[
D_1 \Sigma_1 D_1' = \begin{bmatrix}
1 & \rho_{12} & \ldots & \rho_{1k} \\
\rho_{21} & 1 & \ldots & \rho_{2k} \\
\vdots & \vdots & \ddots & \vdots \\
\rho_{k1} & \rho_{k2} & \ldots & 1
\end{bmatrix},
\tag{8}
\]

\[
D_2 \Sigma_2 D_2' = \frac{1}{2} \begin{bmatrix}
r_1^2 & r_1 r_2 \rho_{12} & \ldots & r_1 r_k \rho_{1k} \\
r_1 r_2 \rho_{12} & r_2^2 & \ldots & r_2 r_k \rho_{2k} \\
\vdots & \vdots & \ddots & \vdots \\
r_1 r_k \rho_{k1} & r_2 r_k \rho_{k2} & \ldots & r_k^2
\end{bmatrix},
\tag{10}
\]

and \(\rho_{ij} = \sigma_{ij}/(\sigma_i \sigma_j)\) is the correlation of the excess returns of portfolios \(i\) and \(j\). Therefore, (6) can be expressed as

\[
\sqrt{n} [u(\hat{\theta}) - u(\theta)] \xrightarrow{D} N(0, \Omega)
\tag{9}
\]

where

\[
\Omega = \frac{1}{2} \begin{bmatrix}
2 + r_1^2 & 2 \rho_{12} + r_1 r_2 \rho_{12} & \ldots & 2 \rho_{1k} + r_1 r_k \rho_{1k} \\
2 \rho_{12} + r_1 r_2 \rho_{12} & 2 + r_2^2 & \ldots & 2 \rho_{2k} + r_2 r_k \rho_{2k} \\
\vdots & \vdots & \ddots & \vdots \\
2 \rho_{k1} + r_1 r_k \rho_{k1} & 2 \rho_{k2} + r_2 r_k \rho_{k2} & \ldots & 2 + r_k^2
\end{bmatrix}.
\tag{10}
\]

Note that the asymptotic variance of \(r_i\) in (9) is \(1 + r_i^2/2\) which is the same as Equation (8) given in Lo (2002). When \(k = 2\), \(Var(r_1 - r_2) = \frac{1}{n}[2 - 2\rho_{12} + \frac{1}{2}(r_1^2 + r_2^2 - 2r_1 r_2 \rho_{12})]\) which is the same as Equation (13) in Memmel (2003).
3 Hypothesis Testing

Having derived the asymptotic distribution of the Sharpe ratio $u(\theta) = (r_1, \ldots, r_k)'$, we are going to test the hypothesis $H_0$ stated in (3). The standard multivariate method known as the repeated Measures Design for comparing Treatments (see for example, Johnson and Wichern 2002, p278) is used to test $H_0$. We first define the $(k-1) \times k$ constant matrix $C$ such that

$$C = \begin{pmatrix} 1 & -1 & 0 & \ldots & 0 \\ 0 & 1 & -1 & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ldots & \ldots & 1 & -1 \end{pmatrix}. \quad (11)$$

The hypothesis $H_0$ in (3) is then equivalent to

$$H_0 : C u(\theta) = 0 \quad \text{versus} \quad H_1 : C u(\theta) \neq 0. \quad (12)$$

Thereafter, the Hotelling’s $T^2$ test

$$T^2 = nk[C u(\hat{\theta})]'(C \hat{\Omega} C')^{-1}[C u(\hat{\theta})] \quad (13)$$

can be applied to test this hypothesis. For the $\alpha$ level, we reject the hypothesis $H_0$ in (12) if

$$T^2 > \frac{(n-1)(k-1)}{(n-k+1)} F_{k-1, n-k+1}(\alpha) \quad (14)$$

and conclude that in $H_1$, $C u(\theta) \neq 0$ where $F_{k-1, n-k+1}(\alpha)$ is the upper $(100\alpha)^{th}$ percentile of an $F$-distribution with $k-1$ and $n-k+1$ degrees of freedom and $\hat{\Omega}$ is the estimate of $\Omega$ in (10) such that the unknown parameters $r_i$ and $\rho_{ij}$ are replaced by their sample estimates.
4 An empirical example

Standard and Poor’s Depository Receipts, (SPDRs or ‘spiders’ with ticker symbol: SPY) began trading in January 1993. They track the S&P 500 Index, and are created and redeemed via ‘creation units’ of 50,000 shares. The acceptance and wide use of SPDRs led to the introduction in March, 1996, of seventeen exchange traded funds (ETFs), now known as iShares which began to trade on the American Stock Exchange and were designed to track the Morgan Stanley Capital International (MSCI) foreign stock market indices. A listing of the country market indices, their ticker symbols and inception dates are presented in Table 1. The iShares are the ‘creation units’ cause their shares to trade in the secondary market just like ordinary shares. These ‘creation units’ contain portfolios of securities designed to represent a particular MSCI Index. An opportunity for applying these innovations emerged with the introduction of country index funds allows investors to trade shares of several well-diversified portfolios representing different countries’ indices to achieve international portfolio diversification which was first advocated by Grubel as early as in 1968.

To illustrate the testing procedure developed in the previous section, we use the daily returns of 17 iShares and the return of the US SPDR (which is treated as the 18th iShare for comparison) from March 19, 1996 to December 31, 2003 as an example. In addition, we use the 3-month T-bill as the risk free return. The daily return of the iShare data were obtained from the Center for Research in Security Prices (CRSP) while the 3-month T-bill were obtained from Datastream. Thereafter, we obtain totally 18 returns computed by

\[ r_i = \frac{\bar{x}_i - r_f}{s_i} \quad \text{for} \quad i = 1, \ldots, 18 \quad (15) \]

where \( \bar{x}_i \) and \( s_i \) is the sample mean and sample standard deviation of the return of the
ith iShare and \( r_f \) is the 3-month T-bill rate. It is useful for investors to test whether the returns of these 18 portfolios have the same sharpe ratio. Rejecting it implies that among these 18 iShares, some iShares outperform other iShares and this information could be useful for investors to make wiser decisions when investing in the iShares. As the Sharpe ratio test developed by Jobson and Korkie (1981), Lo (2002) and Memmel (2003) could only be used to compare the Sharpe ratios of the returns for a pair of portfolios, not for the entire 18 iShares, we will apply the test developed in this paper for this purpose.

We first depict in Table 2 the means, standard deviations and the Sharpe ratios for the returns of these 18 iShares. The Sharpe ratios are then used to rank the iShares from the highest to the lowest ratios. In order to test the performance of the iShares, we first pair up the iShares and apply the pair-wise Sharpe ratio test statistic developed by Jobson and Korkie (1981), Lo (2002) and Memmel (2003) to test their equality. For simplicity, we display the \( p \)-values of these results only for the entire period in Tables 3A and 3B.\(^2\) The results do not allow us to reject the equality of the Sharpe ratios for any pair of the iShares. This may explain why investors conclude that the returns of all the 18 iShares are indistinguishable, including those with the highest and the lowest ratios.

As the pair-wise test does not take into consideration the joint effect of all the portfolios, the conclusion drawn by this pair-wise test could be ambiguous in their evaluation of the iShares. Specifically, ambiguity could present between measure or within measure or both. To circumvent this problem, we apply the test developed in this paper for comparison. Using the notations defined in Section 3, we have \( n = 1961 \) and \( k = 18 \). Table 4 gives the values of the Hotelling’s \( T^2 \) statistic defined in (13) and their corresponding \( p \)-values.

\(^2\)We also conduct the test for the data in each year. As the conclusions drawn from the results in each year are the same as those in the entire period, we only report the results in the entire period. The results in each year are available on request.
for each year from 1996 to 2003 and the \( p \)-value for the entire sample period. The matrix \( D_1 \Sigma_1 D_1' \) in (8) is the correlation matrix of these excess returns. This is an \( 18 \times 18 \) matrix which we skip reporting it in this paper as it is too large to be displayed here. Once this matrix is computed, it is straightforward to compute \( \hat{\Omega} \) from (10) and thereafter calculate the Hotelling’s \( T^2 \) statistic from (13). It turns out that \( T^2 = 88.2277 \) for the entire period, which is much greater than \( [(n-1)(k-1)F_{k-1,n-k+1}(0.01)]/(n-k+1) = 33.843 \). Similarly, the computed Hotelling’s \( T^2 \) statistic for each year is much bigger than the corresponding critical value at the 1% level. Thus, there is sufficient evidence to conclude that these Sharpe ratios are different for each year as well as for the entire sample, contradicting the conclusion drawn by the pair-wise Sharpe ratio test. Nonetheless, different from applying pair-wise Sharpe ratio test, one will conclude that some iShares outperform other iShares in the market when they apply the test developed in this paper. This additional piece of information provided by applying our test could be useful for investors in their investment decisions.
Table 1: Ticker symbols and the date of inception for the 18 iShares

<table>
<thead>
<tr>
<th>Country Fund</th>
<th>Symbol</th>
<th>Inception Date</th>
<th>Country Fund</th>
<th>Symbol</th>
<th>Inception Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>U.S.</td>
<td>SPY</td>
<td>Jan-93</td>
<td>Japan</td>
<td>EWJ</td>
<td>Mar-96</td>
</tr>
<tr>
<td>Australia</td>
<td>EWA</td>
<td>Mar-96</td>
<td>Malaysia</td>
<td>EWM</td>
<td>Mar-96</td>
</tr>
<tr>
<td>Austria</td>
<td>EWO</td>
<td>Mar-96</td>
<td>Mexico</td>
<td>EWW</td>
<td>Mar-96</td>
</tr>
<tr>
<td>Belgium</td>
<td>EWK</td>
<td>Mar-96</td>
<td>Netherlands</td>
<td>EWN</td>
<td>Mar-96</td>
</tr>
<tr>
<td>Canada</td>
<td>EWC</td>
<td>Mar-96</td>
<td>Singapore</td>
<td>EWS</td>
<td>Mar-96</td>
</tr>
<tr>
<td>France</td>
<td>EWQ</td>
<td>Mar-96</td>
<td>Spain</td>
<td>EWP</td>
<td>Mar-96</td>
</tr>
<tr>
<td>Germany</td>
<td>EWG</td>
<td>Mar-96</td>
<td>Sweden</td>
<td>EWD</td>
<td>Mar-96</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>EWH</td>
<td>Mar-96</td>
<td>Switzerland</td>
<td>EWL</td>
<td>Mar-96</td>
</tr>
<tr>
<td>Italy</td>
<td>EWI</td>
<td>Mar-96</td>
<td>United Kingdom</td>
<td>EWU</td>
<td>Mar-96</td>
</tr>
</tbody>
</table>

Note: Information was obtained from Gasbarro et al (2005).

Table 2: Simple Descriptive Statistics and Sharpe Ratios for the 18 iShares

<table>
<thead>
<tr>
<th>Ishare</th>
<th>mean</th>
<th>SD</th>
<th>Sharpe</th>
<th>Ishare</th>
<th>mean</th>
<th>SD</th>
<th>Sharpe</th>
</tr>
</thead>
<tbody>
<tr>
<td>EWA</td>
<td>0.000148</td>
<td>0.015621</td>
<td>0.002471</td>
<td>EWM</td>
<td>-0.00038</td>
<td>0.026585</td>
<td>-0.01827</td>
</tr>
<tr>
<td>EWO</td>
<td>0.000105</td>
<td>0.014993</td>
<td>-0.0003</td>
<td>EWW</td>
<td>0.000254</td>
<td>0.022415</td>
<td>0.006473</td>
</tr>
<tr>
<td>EWK</td>
<td>-6.60E-05</td>
<td>0.02124</td>
<td>-0.00826</td>
<td>EWN</td>
<td>-2.40E-06</td>
<td>0.016401</td>
<td>-0.0068</td>
</tr>
<tr>
<td>EWC</td>
<td>0.000165</td>
<td>0.015855</td>
<td>0.003519</td>
<td>EWS</td>
<td>-0.00037</td>
<td>0.02292</td>
<td>-0.02109</td>
</tr>
<tr>
<td>EWQ</td>
<td>0.000239</td>
<td>0.015871</td>
<td>0.00821</td>
<td>EWP</td>
<td>0.000369</td>
<td>0.016359</td>
<td>0.015912</td>
</tr>
<tr>
<td>EWG</td>
<td>9.50E-05</td>
<td>0.017347</td>
<td>-0.00081</td>
<td>EWD</td>
<td>5.75E-05</td>
<td>0.021419</td>
<td>-0.00241</td>
</tr>
<tr>
<td>EWH</td>
<td>-0.00014</td>
<td>0.022418</td>
<td>-0.01094</td>
<td>EWL</td>
<td>8.41E-05</td>
<td>0.015987</td>
<td>-0.00157</td>
</tr>
<tr>
<td>EWI</td>
<td>0.000193</td>
<td>0.016515</td>
<td>0.005093</td>
<td>EWU</td>
<td>0.000117</td>
<td>0.014988</td>
<td>0.000555</td>
</tr>
<tr>
<td>EWJ</td>
<td>-0.00023</td>
<td>0.017406</td>
<td>-0.01934</td>
<td>SPY</td>
<td>0.000267</td>
<td>0.013134</td>
<td>0.012055</td>
</tr>
</tbody>
</table>
Table 3A: Sharpe Ratio Test Proposed by Memmel 2003

<table>
<thead>
<tr>
<th></th>
<th>EWA</th>
<th>EWO</th>
<th>EWK</th>
<th>EWC</th>
<th>EWQ</th>
<th>EWG</th>
<th>EWH</th>
<th>EWI</th>
<th>EWJ</th>
</tr>
</thead>
<tbody>
<tr>
<td>EWA</td>
<td>1.000</td>
<td>0.921</td>
<td>0.711</td>
<td>0.968</td>
<td>0.820</td>
<td>0.897</td>
<td>0.602</td>
<td>0.920</td>
<td>0.404</td>
</tr>
<tr>
<td>EWO</td>
<td>0.921</td>
<td>1.000</td>
<td>0.775</td>
<td>0.893</td>
<td>0.749</td>
<td>0.985</td>
<td>0.717</td>
<td>0.839</td>
<td>0.505</td>
</tr>
<tr>
<td>EWK</td>
<td>0.711</td>
<td>0.775</td>
<td>1.000</td>
<td>0.674</td>
<td>0.506</td>
<td>0.766</td>
<td>0.924</td>
<td>0.594</td>
<td>0.696</td>
</tr>
<tr>
<td>EWC</td>
<td>0.968</td>
<td>0.893</td>
<td>0.674</td>
<td>1.000</td>
<td>0.843</td>
<td>0.851</td>
<td>0.577</td>
<td>0.948</td>
<td>0.382</td>
</tr>
<tr>
<td>EWQ</td>
<td>0.820</td>
<td>0.749</td>
<td>0.506</td>
<td>0.843</td>
<td>1.000</td>
<td>0.565</td>
<td>0.438</td>
<td>0.865</td>
<td>0.260</td>
</tr>
<tr>
<td>EWG</td>
<td>0.897</td>
<td>0.985</td>
<td>0.766</td>
<td>0.851</td>
<td>0.565</td>
<td>1.000</td>
<td>0.666</td>
<td>0.751</td>
<td>0.441</td>
</tr>
<tr>
<td>EWH</td>
<td>0.602</td>
<td>0.717</td>
<td>0.924</td>
<td>0.577</td>
<td>0.438</td>
<td>0.666</td>
<td>1.000</td>
<td>0.529</td>
<td>0.711</td>
</tr>
<tr>
<td>EWI</td>
<td>0.920</td>
<td>0.839</td>
<td>0.594</td>
<td>0.948</td>
<td>0.865</td>
<td>0.751</td>
<td>0.529</td>
<td>1.000</td>
<td>0.347</td>
</tr>
<tr>
<td>EWJ</td>
<td>0.404</td>
<td>0.505</td>
<td>0.696</td>
<td>0.382</td>
<td>0.260</td>
<td>0.441</td>
<td>0.711</td>
<td>0.347</td>
<td>1.000</td>
</tr>
<tr>
<td>EWM</td>
<td>0.451</td>
<td>0.545</td>
<td>0.739</td>
<td>0.432</td>
<td>0.343</td>
<td>0.530</td>
<td>0.776</td>
<td>0.409</td>
<td>0.969</td>
</tr>
<tr>
<td>EWW</td>
<td>0.877</td>
<td>0.818</td>
<td>0.602</td>
<td>0.905</td>
<td>0.942</td>
<td>0.761</td>
<td>0.458</td>
<td>0.956</td>
<td>0.307</td>
</tr>
<tr>
<td>EWN</td>
<td>0.725</td>
<td>0.811</td>
<td>0.951</td>
<td>0.677</td>
<td>0.405</td>
<td>0.745</td>
<td>0.868</td>
<td>0.543</td>
<td>0.618</td>
</tr>
<tr>
<td>EWS</td>
<td>0.373</td>
<td>0.473</td>
<td>0.652</td>
<td>0.358</td>
<td>0.249</td>
<td>0.414</td>
<td>0.625</td>
<td>0.313</td>
<td>0.942</td>
</tr>
<tr>
<td>EWP</td>
<td>0.617</td>
<td>0.542</td>
<td>0.329</td>
<td>0.616</td>
<td>0.679</td>
<td>0.371</td>
<td>0.294</td>
<td>0.565</td>
<td>0.164</td>
</tr>
<tr>
<td>EWD</td>
<td>0.856</td>
<td>0.939</td>
<td>0.825</td>
<td>0.796</td>
<td>0.610</td>
<td>0.938</td>
<td>0.739</td>
<td>0.733</td>
<td>0.502</td>
</tr>
<tr>
<td>EWL</td>
<td>0.882</td>
<td>0.963</td>
<td>0.788</td>
<td>0.848</td>
<td>0.652</td>
<td>0.972</td>
<td>0.720</td>
<td>0.770</td>
<td>0.491</td>
</tr>
<tr>
<td>EWU</td>
<td>0.942</td>
<td>0.975</td>
<td>0.730</td>
<td>0.905</td>
<td>0.704</td>
<td>0.944</td>
<td>0.632</td>
<td>0.834</td>
<td>0.425</td>
</tr>
<tr>
<td>SPY</td>
<td>0.766</td>
<td>0.696</td>
<td>0.524</td>
<td>0.787</td>
<td>0.904</td>
<td>0.685</td>
<td>0.469</td>
<td>0.828</td>
<td>0.319</td>
</tr>
</tbody>
</table>
Table 3B: Sharpe Ratio Test Proposed by Memmel 2003

<table>
<thead>
<tr>
<th></th>
<th>EWM</th>
<th>EWW</th>
<th>EWN</th>
<th>EWS</th>
<th>EWP</th>
<th>EWD</th>
<th>EWL</th>
<th>EWU</th>
<th>SPY</th>
</tr>
</thead>
<tbody>
<tr>
<td>EWA</td>
<td>0.451</td>
<td>0.877</td>
<td>0.725</td>
<td>0.373</td>
<td>0.617</td>
<td>0.856</td>
<td>0.882</td>
<td>0.942</td>
<td>0.766</td>
</tr>
<tr>
<td>EWO</td>
<td>0.545</td>
<td>0.818</td>
<td>0.811</td>
<td>0.473</td>
<td>0.542</td>
<td>0.939</td>
<td>0.963</td>
<td>0.975</td>
<td>0.696</td>
</tr>
<tr>
<td>EWK</td>
<td>0.739</td>
<td>0.602</td>
<td>0.951</td>
<td>0.652</td>
<td>0.329</td>
<td>0.825</td>
<td>0.788</td>
<td>0.730</td>
<td>0.524</td>
</tr>
<tr>
<td>EWC</td>
<td>0.432</td>
<td>0.905</td>
<td>0.677</td>
<td>0.358</td>
<td>0.616</td>
<td>0.796</td>
<td>0.848</td>
<td>0.905</td>
<td>0.787</td>
</tr>
<tr>
<td>EWQ</td>
<td>0.343</td>
<td>0.942</td>
<td>0.405</td>
<td>0.249</td>
<td>0.679</td>
<td>0.610</td>
<td>0.652</td>
<td>0.704</td>
<td>0.904</td>
</tr>
<tr>
<td>EWG</td>
<td>0.530</td>
<td>0.761</td>
<td>0.745</td>
<td>0.414</td>
<td>0.371</td>
<td>0.938</td>
<td>0.972</td>
<td>0.944</td>
<td>0.685</td>
</tr>
<tr>
<td>EWH</td>
<td>0.776</td>
<td>0.458</td>
<td>0.868</td>
<td>0.625</td>
<td>0.294</td>
<td>0.739</td>
<td>0.720</td>
<td>0.632</td>
<td>0.469</td>
</tr>
<tr>
<td>EWI</td>
<td>0.409</td>
<td>0.956</td>
<td>0.543</td>
<td>0.313</td>
<td>0.565</td>
<td>0.733</td>
<td>0.770</td>
<td>0.834</td>
<td>0.828</td>
</tr>
<tr>
<td>EWJ</td>
<td>0.969</td>
<td>0.307</td>
<td>0.618</td>
<td>0.942</td>
<td>0.164</td>
<td>0.502</td>
<td>0.491</td>
<td>0.425</td>
<td>0.319</td>
</tr>
<tr>
<td>EWM</td>
<td>1.000</td>
<td>0.363</td>
<td>0.686</td>
<td>0.909</td>
<td>0.236</td>
<td>0.586</td>
<td>0.570</td>
<td>0.504</td>
<td>0.342</td>
</tr>
<tr>
<td>EWW</td>
<td>0.363</td>
<td>1.000</td>
<td>0.591</td>
<td>0.266</td>
<td>0.703</td>
<td>0.723</td>
<td>0.757</td>
<td>0.810</td>
<td>0.859</td>
</tr>
<tr>
<td>EWN</td>
<td>0.686</td>
<td>0.591</td>
<td>1.000</td>
<td>0.579</td>
<td>0.229</td>
<td>0.833</td>
<td>0.804</td>
<td>0.723</td>
<td>0.557</td>
</tr>
<tr>
<td>EWS</td>
<td>0.909</td>
<td>0.266</td>
<td>0.579</td>
<td>1.000</td>
<td>0.150</td>
<td>0.471</td>
<td>0.460</td>
<td>0.389</td>
<td>0.301</td>
</tr>
<tr>
<td>EWP</td>
<td>0.236</td>
<td>0.703</td>
<td>0.229</td>
<td>0.150</td>
<td>1.000</td>
<td>0.383</td>
<td>0.416</td>
<td>0.468</td>
<td>0.903</td>
</tr>
<tr>
<td>EWD</td>
<td>0.586</td>
<td>0.723</td>
<td>0.833</td>
<td>0.471</td>
<td>0.383</td>
<td>1.000</td>
<td>0.971</td>
<td>0.897</td>
<td>0.648</td>
</tr>
<tr>
<td>EWL</td>
<td>0.570</td>
<td>0.757</td>
<td>0.804</td>
<td>0.460</td>
<td>0.416</td>
<td>0.971</td>
<td>1.000</td>
<td>0.922</td>
<td>0.667</td>
</tr>
<tr>
<td>EWU</td>
<td>0.504</td>
<td>0.810</td>
<td>0.723</td>
<td>0.389</td>
<td>0.468</td>
<td>0.897</td>
<td>0.922</td>
<td>1.000</td>
<td>0.718</td>
</tr>
<tr>
<td>SPY</td>
<td>0.342</td>
<td>0.859</td>
<td>0.557</td>
<td>0.301</td>
<td>0.903</td>
<td>0.648</td>
<td>0.667</td>
<td>0.718</td>
<td>1.000</td>
</tr>
</tbody>
</table>
Table 4: Hotelling's statistic and the corresponding P values

<table>
<thead>
<tr>
<th>Year</th>
<th>Hotelling's statistic</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996</td>
<td>196.3010</td>
<td>0</td>
</tr>
<tr>
<td>1997</td>
<td>374.7384</td>
<td>0</td>
</tr>
<tr>
<td>1998</td>
<td>168.3597</td>
<td>0</td>
</tr>
<tr>
<td>1999</td>
<td>306.5009</td>
<td>0</td>
</tr>
<tr>
<td>2000</td>
<td>119.6312</td>
<td>0</td>
</tr>
<tr>
<td>2001</td>
<td>118.7369</td>
<td>0</td>
</tr>
<tr>
<td>2002</td>
<td>74.6715</td>
<td>0</td>
</tr>
<tr>
<td>2003</td>
<td>128.9687</td>
<td>0</td>
</tr>
<tr>
<td>Entire sample</td>
<td>88.2227</td>
<td>0</td>
</tr>
</tbody>
</table>

5 Conclusion

Adopting the repeated measures design, this paper extends the work of Jobson and Korkie (1981), Lo (2002) and Memmel (2003) by developing the multivariate test statistic to test the hypothesis of the equality of multiple Sharpe ratios. In addition, we develop the asymptotic distribution of the statistics, and provide a means of correcting the bias in Sharpe Ratio noted by Miller and Gehr (1978). However, this bias will decreases as the sample size increases. Thus, this bias will not affect the testing procedure introduced in our paper asymptotically.

Achour, et al (1984) develop a test for multiple Sharpe Ratio based on the sum of pair-wise difference of two portfolios. They propose two asymptotic tests using normal distribution and chi-square distribution. Their approach can be considered as a special case of the multivariate test developed in this paper. Our approach is more general and
uses the well-known Hotelling’s $T^2$ test in multivariate analysis.

To illustrate the applicability of the multivariate test statistic developed in our paper, we apply the statistic to test the equality of Sharpe ratios for the eighteen iShares. The results could then be used to evaluate the performance of the iShares. We note that the multivariate test statistic developed in our paper could also be used in many different contexts in Finance and Economics. For example, the statistic could be used to extend the work of Levy (1972), Cumby and Glen (1990), Jorion (1991), Grinblatt and Titman (1994) and MacKinlay and Pastor (2000) in the evaluation of multiple portfolio performance, which, in turn, could be used to test market efficiency. The results will be useful in risk management and investment decision making.

Further extension includes applying our approach to modify several other existing pair-wise test statistics into multivariate test statistics. For example, Cerny (2003) extends Sharpe’s work to the Generalized Sharpe Ratios which provide a consistent ranking of investment opportunities even when asset returns are highly non-normal. Favre and Galeano’s (2002) develop the Modified Value-at-Risk (VaR) which replaces the standard deviation in the Sharpe Ratio to correctly assess non-normal returns. Chen (2005) considers nonparametric estimation of VaR and associated standard error estimation for dependent financial returns. Kuester, Mittnik, and Paolella (2006) compare the out-of-sample performance of existing methods and some new models for predicting value-at-risk (VaR) in a univariate context. They find that most approaches perform inadequately and find that a hybrid method, combining a heavy-tailed generalized autoregressive conditionally heteroskedastic (GARCH) filter with an extreme value theory-based approach, performs best overall. One could easily apply our approach to extend the above work into the corresponding multivariate test statistic which could shed new light on asset investments.
References


