DISCUSSION PAPERS IN STATISTICS AND ECONOMETRICS

SEMINAR OF ECONOMIC AND SOCIAL STATISTICS
UNIVERSITY OF COLOGNE

No. 7/07

Testing for the Best Alternative with an Application to Performance Measurement

by

Gabriel Frahm

4th version
September 3, 2007

DISKUSSIONSBEITRÄGE ZUR STATISTIK UND ÖKONOMETRIE

SEMINAR FÜR WIRTSCHAFTS- UND SOZIALSTATISTIK
UNIVERSITÄT ZU KÖLN

Albertus-Magnus-Platz, D-50923 Köln, Deutschland

This page intentionally left blank.

DISCUSSION PAPERS IN STATISTICS AND ECONOMETRICS

SEMINAR OF ECONOMIC AND SOCIAL STATISTICS
UNIVERSITY OF COLOGNE

No. 7/07

Testing for the Best Alternative with an Application to Performance Measurement

by

Gabriel Frahm[1]

4th version
September 3, 2007

Abstract

Suppose that we are searching for the maximum of many unknown and analytically untractable quantities or, say, the 'best alternative' among several candidates. If our decision is based on historical or simulated data there is some sort of selection bias and it is not evident if our choice is significantly better than any other. In the present work a large sample test for the best alternative is derived in a rather general setting. The test is demonstrated by an application to financial data and compared with the Jobson-Korkie test for the Sharpe ratios of two asset portfolios. We find that ignoring conditional heteroscedasticity and non-normality of asset returns can lead to misleading decisions. In contrast, the presented test for the best alternative accounts for these kinds of phenomena.

Keywords: Ergodicity, Gordin's condition, heteroscedasticity, Jobson-Korkie test, Monte Carlo simulation, performance measurement, Sharpe ratio.

AMS Subject Classification: Primary 62G10, Secondary 91B28.

[1]Department of Economic and Social Statistics, Statistics & Econometrics, University of Cologne, Albertus-Magnus-Platz, D-50923 Cologne, Germany.

This page intentionally left blank.

TESTING FOR THE BEST ALTERNATIVE WITH AN APPLICATION TO PERFORMANCE MEASUREMENT

GABRIEL FRAHM
UNIVERSITY OF COLOGNE
STATISTICS AND ECONOMETRICS
MEISTER-EKKEHART-STR. 9
50937 COLOGNE
GERMANY

Abstract. Suppose that we are searching for the maximum of many unknown and analytically untractable quantities or, say, the 'best alternative' among several candidates. If our decision is based on historical or simulated data there is some sort of selection bias and it is not evident if our choice is significantly better than any other. In the present work a large sample test for the best alternative is derived in a rather general setting. The test is demonstrated by an application to financial data and compared with the Jobson-Korkie test for the Sharpe ratios of two asset portfolios. We find that ignoring conditional heteroscedasticity and non-normality of asset returns can lead to misleading decisions. In contrast, the presented test for the best alternative accounts for these kinds of phenomena.

Keywords: Ergodicity, Gordin's condition, heteroscedasticity, Jobson-Korkie test, Monte Carlo simulation, performance measurement, Sharpe ratio.

AMS Subject Classification: Primary 62G10, Secondary 91B28.

Motivation

In many practical situations we cannot calculate some number analytically. Then it is often possible to use Monte Carlo simulation for approximating the desired quantity. Standard large sample theory can be applied for controlling such kind of approximations. Now suppose that we are searching for the maximum of some unknown and analytically untractable quantities. Thus we could choose the largest outcome given by Monte Carlo simulation. However, since we take the best result from a set of given outcomes there is some sort of selection bias and it is not evident if our choice is *significantly* better or at least not worse than any other. The same problem frequently occurs in statistical inference or decisions under uncertainty when searching for the 'best alternative' such as portfolio optimization. In the following I will derive a large sample test for the best alternative in a rather general setting. The presented test is demonstrated by an application to financial data. It is shown that the Jobson-Korkie test for the Sharpe ratios of two asset portfolios can be generalized to ergodic stationary stochastic processes satisfying Gordin's condition. The resulting test for the best alternative accounts for conditional heteroscedasticity and non-normality of asset returns in contrast to the Jobson-Korkie test.

1. Hypothesis Test for the Best Alternative

1.1. Basic Assumptions and Notation.

Let $\mu = (\mu_1, \ldots, \mu_d) \in \mathbb{R}^d$ be an unknown vector of quantities and we are searching for the *best alternative*

$$i^* := \arg\max_j \big\{ \mu_j \colon j = 1, \ldots, d \big\} .$$

It is worth to mention that i^* does not need to be unique. That means there can be several equivalent and optimal alternatives. In contrast, let $i \in \{1, \ldots, d\}$ be our specific *choice*, i.e. we believe that there is no other alternative better than μ_i. We will set $i = 1$ for notational convenience and without loss of generality. Hence, we want to support the alternative hypothesis

$$H_1 : \mu_1 \geq \mu_2, \ldots, \mu_d$$

vs. the null hypothesis $H_0 : \neg H_1$. If we can reject H_0, our choice turns out to be significantly optimal among all given alternatives.

Let (X_n) be a sequence of d-dimensional random vectors such that

$$a_n\big(X_n - \mu\big) \overset{\mathrm{d}}{\longrightarrow} \xi\,, \qquad n \longrightarrow \infty\,,$$

where (a_n) is some sequence of real numbers growing to infinity and ξ is a d-dimensional random vector. It is supposed that the cumulative distribution function (c.d.f.) of ξ does not depend on μ. By Cramér's theorem (Davidson, 1994, p. 355) it follows that $X_n \rightarrow_{\mathrm{p}} \mu$ as $n \rightarrow \infty$. Hence, we can think of X_n as a convenient approximation of μ if n is large. Due to the Central limit theorem (CLT) we will typically encounter $a_n = \sqrt{n}$ and ξ has a multivariate normal distribution with zero mean and covariance matrix Σ.

1.2. **Test Procedure.** A crucial point of the following test is that i must be fixed *without* examining X_n or say, more precisely, the choice must not depend on the data which are used for testing the aforementioned hypothesis. Otherwise the presented method would suffer from a selection bias. Indeed, this is not a serious drawback of the procedure. For instance, consider a Monte Carlo simulation. In that case we can simply run the process (X_n) a first time so as to choose the largest component of X_n, that is

$$i = \arg\max_j \big\{X_{jn} : \; j = 1, \ldots, d\big\}\,.$$

After that we start a new run of (X_n) and apply the following test with respect to the choice made by the first run. In case of historical data we can simply divide the overall sample into two sub-samples, i.e. a calibration and a validation sample. Then the choice can be made by using the calibration sample, whereas the test has to be applied to the validation sample.

Define the $(d-1) \times d$ matrix

$$\Delta := \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & -1 \end{bmatrix}$$

and note that due to the Continuous mapping theorem (Davidson, 1994, p. 355) we obtain

$$a_n\big(\Delta X_n - \Delta\mu\big) \overset{\mathrm{d}}{\longrightarrow} \Delta\xi\,, \qquad n \longrightarrow \infty\,.$$

Now the alternative hypothesis can be compactly written as $H_1 : \Delta\mu \geq 0$. In case $d = 2$ we will obtain a simple Gauss-type test for the null hypothesis $H_{02} : \mu_1 < \mu_2$. In the general multivariate case the global hypothesis H_1 can be supported whenever $H_{1j} : \mu_1 \geq \mu_j$ survives after each comparison with $j = 2, \ldots, d$. This is an important implication of the following theorem.

Theorem 1.2.1. *Let $\zeta = (\zeta_1, \ldots, \zeta_k)$ be a random vector and consider $Z = \eta + \zeta$ where $\eta \in \mathbb{R}^k$ but not $\eta \geq 0$. Let λ_j be the β-quantile of ζ_j for $j = 1, \ldots, k$ and $0 < \beta < 1$. Then $\mathbb{P}\big(Z > \lambda\big) \leq 1 - \beta$ with $\lambda = (\lambda_1, \ldots, \lambda_k) \in \mathbb{R}^k$.*

Proof. At least one component of η must be negative, say $\eta_j < 0$. Now the assertion follows immediately by noting that $\mathbb{P}(Z > \lambda) \leq \mathbb{P}(Z_j > \lambda_j) \leq 1 - \beta$. □

In our case η represents $\Delta\mu$, $k = d-1$, $\beta = 1-\alpha$ with $0 < \alpha < 1$, $\zeta = \Delta\xi/a_n$, and $Z = \Delta X_n$. Hence, we can reject H_0 if $\Delta X_n > \lambda$ or, following the usual notation of large sample theory, $T := a_n \Delta X_n > \tau$, where $\tau = (\tau_1, \ldots, \tau_{d-1}) := a_n\lambda$. The $(d-1)\times 1$ vector τ contains the $(1-\alpha)$-quantiles of $\Delta\xi$. Theorem 1.2.1 guarantees that our choice is significantly optimal among all given alternatives whenever it is significantly better or not worse than every other candidate on the *same* level α. That means if each pairwise test $H_0: \mu_1 < \mu_j$ vs. $H_1: \mu_1 \geq \mu_j$ possesses a significance level of α then the overall test $H_1\colon \mu_1 \geq \mu_2, \ldots, \mu_d$ vs. $H_0\colon \neg H_1$ works on the same level.

In many practical situations we do not know the exact c.d.f. of $\Delta\xi$. However, we can often calculate or simulate the c.d.f. of ξ_n, where (ξ_n) is some sequence of d-dimensional random vectors such that $\xi_n \to_{\mathrm{d}} \xi$ as $n \to \infty$. This can be used for a large sample approximation of the critical thresholds $\tau_1, \ldots, \tau_{d-1}$. For instance, suppose that $X_1, \ldots, X_n$ is a sample of independent copies of a random vector X with mean vector μ and positive definite covariance matrix Σ. We assume that μ and Σ are unknown. From the CLT we know that

$$\sqrt{n}\cdot\left(\frac{1}{n}\cdot\sum_{i=1}^{n} X_i - \mu\right) \xrightarrow{\mathrm{d}} \mathcal{N}\big(0,\Sigma\big)\,, \qquad n \longrightarrow \infty\,.$$

For brevity we may denote the sample mean vector by $\overline{X}_n = \big(\overline{X}_{1n}, \ldots, \overline{X}_{dn}\big)$. Now we try to reject $H_{0j}\colon \mu_1 < \mu_j$ by applying the one-sided Gauss test

$$T_{j-1} := \sqrt{n}\cdot\big(\overline{X}_{1n} - \overline{X}_{jn}\big) > \tau_{j-1}\,,$$

for $j = 2, \ldots, d$, where

$$\tau_{j-1} := \sqrt{\sigma_1^2 + \sigma_j^2 - 2\sigma_{1j}}\cdot\Phi^{-1}(1-\alpha)\,.$$

Here σ_j^2 represents the variance of the jth component of X ($j = 1, \ldots, d$), σ_{1j} is the covariance between its first and jth component ($j = 2, \ldots, d$), and Φ^{-1} denotes the quantile function of the standard normal distribution. Note that the parameters of Σ are unknown but we can substitute Σ by the sample covariance matrix

$$\widehat{\Sigma}_n = \frac{1}{n}\cdot\sum_{i=1}^{n}\big(X_i - \overline{X}_n\big)\big(X_i - \overline{X}_n\big)'$$

because – due to the i.i.d. assumption – the sample covariance matrix is strongly consistent for Σ. Hence, by the Cramér-Wold device (Davidson, 1994, p. 405) it follows that

$$\widehat{\Sigma}_n^{\frac{1}{2}} Y \xrightarrow{\mathrm{d}} \Sigma^{\frac{1}{2}} Y \sim \mathcal{N}(0,\Sigma)\,, \qquad n \longrightarrow \infty\,,$$

where $Y \sim \mathcal{N}(0, I_d)$ and $\Sigma^{\frac{1}{2}}$ denotes a $d\times d$ matrix such that $\Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}\prime} = \Sigma$. That means the critical thresholds $\tau_1, \ldots, \tau_{d-1}$ can be readily approximated by using the sample variances and covariances and we obtain the usual one-sided Gauss test for a joint sample, viz

$$\frac{\overline{X}_{1n} - \overline{X}_{jn}}{\sqrt{\big(\hat\sigma_1^2 + \hat\sigma_j^2 - 2\hat\sigma_{1j}\big)/n}} > \Phi^{-1}(1-\alpha)\,.$$

If this inequality is satisfied for every $j = 2, \ldots, d$, the first alternative is significantly optimal among all given alternatives.

2. Application to Financial Data

2.1. **General Conditions.** Let $P_t \overset{\text{a.s.}}{>} 0$ be the price of an asset at time $t \in \mathbb{Z}$ so that $R_t := P_t/P_{t-1} - 1$ represents the corresponding asset return from $t-1$ to t. It is assumed that (R_t) is strongly stationary and ergodic with $\mathbb{E}(R_t) = \eta$ and $\mathbb{V}\mathrm{ar}(R_t) = \sigma^2 < \infty$. Ergodicity means that any existing and finite moment of R_t

can be consistently estimated by the corresponding sample moment of (R_t). This is guaranteed if $(R_t, \ldots, R_{t+k})$ is asymptotically independent of $(R_{t-n}, \ldots, R_{t-n+l})$ as $n \to \infty$ for all $k, l \in \mathbb{N}$, whilst the components of the considered random vectors generally depend on each other (Hayashi, 2000, p. 101). For the CLT we need some additional restrictions. More precisely, the CLT holds for the sample mean of (R_t) if the centered process $(R_t - \eta)$ satisfies *Gordin's condition*. Let $\mathcal{H}_t := (R_t, R_{t-1}, \ldots)$ be the history of (R_t) at time $t \in \mathbb{Z}$. Roughly speaking, Gordin's condition implies that the impact of $\mathcal{H}_{t-n}$ on the conditional expectation of R_t vanishes as $n \to \infty$ and also that the conditional expectations of R_t do not vary too much in time (Hayashi, 2000, p. 403). In that case it is guaranteed that the CLT holds with an asymptotic or, say, *long-run variance* $\sigma_{\mathrm{L}}^2 := \sum_{k=-\infty}^{\infty} \gamma(k)$ (Hayashi, 2000, p. 401), where γ is the autocovariance function of (R_t). This can be easily extended to any d-dimensional stochastic process (Hayashi, 2000, p. 405) and applied to a broad class of standard time series models. There exist several alternative criteria for the CLT in the context of time series analysis which can be found, e.g., in Brockwell and Davis (1991, p. 213) and Hamilton (1994, p. 195). However, to my knowledge Gordin's condition represents the most unrestrictive set of assumptions concerning the serial dependence structure of a stochastic process (Eagleson, 1975).

It is worth to note that the number of dimensions d is supposed to be fixed or at least $n, d \to \infty$ such that $n/d \to \infty$. If n/d tends to a finite number, the CLT may become invalid and other interesting issues arise from *Random matrix theory* (Bai, 1999). However, if the number of observations relative to the number of assets is large enough, the sample mean is approximately normally distributed under the aforementioned conditions. We additionally assume that the asset return R_t possesses a finite fourth moment and that Gordin's condition is satisfied not only for $(R_t - \eta)$ but also for $\{(R_t - \eta)^2 - \sigma^2\}$. Consider the random variable $X := R/\sigma$ and suppose that the risk-free interest rate is constant and zero without loss of generality. The *Sharpe ratio* $\mu := \eta/\sigma$ (see, e.g., Campbell et al., 1997, p. 188) is frequently used as a performance measure in finance literature.

2.2. **Asymptotic Distributions.** Concerning the sample mean $\hat{\eta}$ we obtain

$$\sqrt{n} \cdot (\hat{\eta} - \eta) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \sigma_{\mathrm{L}}^2), \qquad n \longrightarrow \infty.$$

The sample variance $\hat{\sigma}^2$ represents a consistent estimator for the stationary variance σ^2 but for estimating the long-run variance σ_{L}^2 we need to estimate the autocovariance function γ of (R_t). Actually, there are many alternative estimation procedures for long-run variances and covariances (see, e.g., Ogaki et al., 2007, Ch. 6). This is not the primary concern of the present work and for the sake of simplicity we can choose a simple box-kernel type estimator, viz

$$\hat{\sigma}_{\mathrm{L}}^2 := \hat{\sigma}^2 + 2 \sum_{k=1}^{l} \hat{\gamma}(k),$$

where $\hat{\gamma}$ is the sample autocovariance function of (R_t) (Hayashi, 2000, p. 142) and $l < n$. However, many empirical studies confirm that $\gamma(k) \approx \hat{\gamma}(k) \approx 0$ for $k \neq 0$ and so we can expect that $\hat{\sigma}_{\mathrm{L}}^2 \approx \hat{\sigma}^2$. The standard error of $\hat{\eta}$ is given by $\epsilon(\hat{\eta}) = \sigma_{\mathrm{L}}/\sqrt{n}$ and this can be estimated by $\hat{\epsilon}(\hat{\eta}) = \hat{\sigma}_{\mathrm{L}}/\sqrt{n}$.

Since $\{(R_t - \eta)^2 - \sigma^2\}$ satisfies Gordin's condition, the sample variance $\hat{\sigma}^2$ is also asymptotically normally distributed, viz

$$\sqrt{n} \cdot (\hat{\sigma}^2 - \sigma^2) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \upsilon_{\mathrm{L}}), \qquad n \longrightarrow \infty.$$

The long-run variance υ_{L} of the squared centered asset returns can be estimated by

$$\hat{\upsilon}_{\mathrm{L}} := \hat{\kappa}(0) + 2 \sum_{k=1}^{l} \hat{\kappa}(k),$$

	Canada	France	Germany	Italy	Japan	UK	USA
$\hat{\sigma}_{\mathrm{L}}^2/\hat{\sigma}^2$	1.1334	1.3834	1.2356	1.9596	2.1995	0.9883	1.0505
$\hat{\upsilon}_{\mathrm{L}}/\hat{\kappa}(0)$	2.1004	1.8611	2.3553	1.8195	2.0844	2.5268	2.0429

TABLE 1. Estimated long-run variances divided by sample variances.

where $\hat{\kappa}$ denotes the sample autocovariance function of $\{(R_t - \eta)^2\}$. Typically, asset returns are conditionally heteroscedastic and thus υ_{L} can become relatively large. This is also confirmed by the following empirical study. We consider monthly excess returns of the MSCI stock indices for the G7 countries Canada, France, Germany, Italy, Japan, UK and USA from January 1970 to September 2006. The sample size corresponds to $n = 456$ and the risk-free interest rate is calculated by the secondary market 3-month US treasury bill rate. Further, the considered indices are adjusted by dividends, splits, etc. and are calculated on the basis of USD stock prices.

For estimating the long-run variances we have to choose an appropriate lag length $l \in \mathbb{N}$. Figure 1 shows the empirical autocorrelations for the squared centered excess returns of the MSCI indices and the *equally weighted portfolio* (EWP) up to $l = 12$. The Ljung-Box test leads to a rejection of the null hypothesis $H_0 : \rho(1) = \ldots = \rho(12) = 0$ in every case except for the EWP, France, and Italy. That means there is a strong evidence of conditional heteroscedasticity for monthly asset returns and we may choose $l = 12$ as an appropriate lag length. Now, Table 1 contains the estimated long-run variances divided by the corresponding sample variances. In most cases the long-run variances of the asset returns roughly correspond to the stationary variances, whereas the long-run variances of the squared asset returns are quite twice as large as the stationary ones. Hence, it is not appropriate to ignore the effect of heteroscedasticity when analyzing the volatility of monthly asset returns.

By applying the well-known 'delta method' we obtain

$$\sqrt{n} \cdot (\hat{\sigma} - \sigma) \xrightarrow{\mathrm{d}} \mathcal{N}\left(0, \frac{\upsilon_{\mathrm{L}}}{4\sigma^2}\right), \qquad n \longrightarrow \infty .$$

The standard error of $\hat{\sigma}$ is given by

$$\epsilon(\hat{\sigma}) := \frac{\sqrt{\upsilon_{\mathrm{L}}/n}}{2\sigma}$$

and its estimator can be denoted by $\hat{\epsilon}(\hat{\sigma}) := \sqrt{\hat{\upsilon}_{\mathrm{L}}/n}/2\hat{\sigma}$.

The Sharpe ratio can be estimated by $\hat{\mu} := \hat{\eta}/\hat{\sigma}$ which is also asymptotically normally distributed since

$$\sqrt{n} \cdot \left(\begin{bmatrix} \hat{\eta} \\ \hat{\sigma}^2 \end{bmatrix} - \begin{bmatrix} \eta \\ \sigma^2 \end{bmatrix} \right) \xrightarrow{\mathrm{d}} \mathcal{N}\left(0, \begin{bmatrix} \sigma_{\mathrm{L}}^2 & \varrho_{\mathrm{L}} \\ \varrho_{\mathrm{L}} & \upsilon_{\mathrm{L}} \end{bmatrix}\right), \qquad n \longrightarrow \infty ,$$

where ϱ_{L} represents the long-run covariance between R_t and $(R_t - \eta)^2$. After applying once again the delta method we obtain

$$\sqrt{n} \cdot (\hat{\mu} - \mu) \xrightarrow{\mathrm{d}} \mathcal{N}\left(0, \frac{\sigma_{\mathrm{L}}^2}{\sigma^2} - \frac{\mu \varrho_{\mathrm{L}}}{\sigma^3} + \frac{\mu^2 \upsilon_{\mathrm{L}}}{4\sigma^4}\right), \qquad n \longrightarrow \infty ,$$

and the standard error of $\hat{\mu}$ can be estimated in the same manner as $\epsilon(\hat{\eta})$ or $\epsilon(\hat{\sigma})$. Schmid and Schmidt (2007) obtain the same asymptotic variance under the assumption of an 'α-mixing process'. As already mentioned this assumption is somewhat more restrictive than Gordin's condition. Schmid and Schmidt (2007) also provide closed-form expressions for the asymptotic variance of the Sharpe ratio in case of a stochastic volatility and a GARCH model.

Table 2 contains the estimated means, standard deviations, and Sharpe ratios for the monthly excess returns of the G7 MSCI indices and the EWP. The corresponding standard error estimates $\hat{\epsilon}(\hat{\eta})$, $\hat{\epsilon}(\hat{\sigma})$, and $\hat{\epsilon}(\hat{\mu})$ are given in the parentheses.

	EWP	Canada	France	Germany	Italy	Japan	UK	USA
$\hat{\eta}$	.0051 (.0026)	.0048 (.0027)	.0062 (.0035)	.0053 (.0031)	.0036 (.0047)	.0058 (.0044)	.0061 (.0030)	.0042 (.0021)
$\hat{\sigma}$	.0437 (.0025)	.0545 (.0038)	.0640 (.0040)	.0603 (.0042)	.0718 (.0039)	.0633 (.0035)	.0638 (.0091)	.0436 (.0029)
$\hat{\mu}$	.1177 (.0620)	.0886 (.0509)	.0967 (.0550)	.0880 (.0523)	.0507 (.0646)	.0909 (.0696)	.0955 (.0479)	.0959 (.0503)

TABLE 2. Means, standard deviations, and Sharpe ratios for the monthly excess returns of the G7 MSCI indices and the EWP.

Obviously, the standard errors for the Sharpe ratios are big despite of the large number of observations. This is a common problem in performance measurement. Now we want to derive an appropriate hypothesis test for the best alternative, i.e. the best performing asset. Without any previous look at the data we may expect that the EWP possesses the largest Sharpe ratio due to the effect of *international diversification* (see, e.g., Jorion, 1985). That means the variance of the EWP return should be relatively small. Indeed, this can be verified in Table 2. Hence, the EWP may serve as the benchmark portfolio and we want to know if its estimated Sharpe ratio $\hat{\mu}_1 = 0.1177$ is *significantly* larger (or at least not smaller) than any other Sharpe ratio.

Also the 2-dimensional random vector $(\hat{\mu}_1, \hat{\mu}_j)$ $(j = 2, \ldots, d)$ is asymptotically normally distributed, i.e.

$$\sqrt{n} \cdot \left(\begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_j \end{bmatrix} - \begin{bmatrix} \mu_1 \\ \mu_j \end{bmatrix} \right) \xrightarrow{\mathrm{d}} \mathcal{N}\left(0, \begin{bmatrix} \vartheta_1^2 & \vartheta_{1j} \\ \vartheta_{j1} & \vartheta_j^2 \end{bmatrix}\right), \qquad n \longrightarrow \infty .$$

After some calculation we obtain

$$\vartheta_{1j} = \frac{\omega_{\mathrm{L}1j}}{\sigma_1 \sigma_j} - \frac{\mu_j \sigma_1 \omega_{\mathrm{L}2j} + \mu_1 \sigma_j \omega_{\mathrm{L}3j}}{2\sigma_1^2 \sigma_j^2} + \frac{\mu_1 \mu_j \omega_{\mathrm{L}4j}}{4\sigma_1^2 \sigma_j^2}$$

for $j = 2, \ldots, d$. Here $\omega_{\mathrm{L}1j}$ represents the long-run covariance between R_{1t} and R_{jt}, $\omega_{\mathrm{L}2j}$ is the long-run covariance between R_{1t} and $(R_{jt} - \eta_j)^2$, $\omega_{\mathrm{L}3j}$ is the long-run covariance between $(R_{1t} - \eta_1)^2$ and R_{jt}, whereas $\omega_{\mathrm{L}4j}$ is the long-run covariance between $(R_{1t} - \eta_1)^2$ and $(R_{jt} - \eta_j)^2$. Now it follows that

$$\sqrt{n} \cdot \left\{ (\hat{\mu}_1 - \hat{\mu}_j) - (\mu_1 - \mu_j) \right\} \xrightarrow{\mathrm{d}} \mathcal{N}\left(0, \vartheta_1^2 + \vartheta_j^2 - 2\vartheta_{1j}\right), \qquad n \longrightarrow \infty .$$

Table 3 contains the values of the test statistic, i.e. $T_{j-1} = \sqrt{n} \cdot (\hat{\mu}_1 - \hat{\mu}_j)$ for $j = 2, \ldots, 8$, the standard errors calculated on the basis of the long-run variances and covariances, and the corresponding 'p-values'. There exists no country with a Sharpe ratio being significantly smaller than the Sharpe ratio of the EWP.

The Jobson-Korkie test (Jobson and Korkie, 1981, Memmel, 2003) is frequently used in the finance literature for comparing the Sharpe ratios of two asset portfolios. For applying this test we have to assume that the asset returns are serially independent and multivariate normally distributed. In that case there is no need to distinguish between long-run, stationary, and conditional variances and covariances of asset returns since these quantities simply coincide. That means $\sigma_{\mathrm{L}1}^2 = \sigma_1^2$, $\sigma_{\mathrm{L}j}^2 = \sigma_j^2$, and $\omega_{\mathrm{L}1j} = \sigma_{1j}$ $(j = 2, \ldots, d)$. Further, by applying some standard results of multivariate analysis (see, e.g., Muirhead, 1982, p. 43) we obtain $\varrho_{\mathrm{L}1} = \varrho_{\mathrm{L}j} = 0$, $\upsilon_{\mathrm{L}1} = 2\sigma_1^4$, $\upsilon_{\mathrm{L}j} = 2\sigma_j^4$, $\omega_{\mathrm{L}2j} = \omega_{\mathrm{L}3j} = 0$, and $\omega_{\mathrm{L}4j} = 2\sigma_{1j}^2$ $(j = 2, \ldots, d)$ so that

$$\sqrt{n} \cdot \left((\hat{\mu}_1 - \hat{\mu}_j) - (\mu_1 - \mu_j) \right) \xrightarrow{\mathrm{d}} \mathcal{N}\left(0, 2\,(1 - \rho_{1j}) + \frac{\mu_1^2 + \mu_j^2 - 2\mu_1 \mu_j \rho_{1j}^2}{2}\right)$$

as $n \to \infty$, where $\rho_{1j} := \sigma_{1j}/(\sigma_1 \sigma_j)$ for $j = 2, \ldots, d$. The latter expression for the asymptotic variance can be found also in Memmel (2003).

	Canada	France	Germany	Italy	Japan	UK	USA
T	.6214 (.9305)	.4478 (.5473)	.6355 (.7453)	1.4320 (.9452)	0.5729 (1.0180)	.4739 (.7208)	.4661 (.8499)
p	.2521	.2066	.1969	.0649	.2868	.2554	.2917

Table 3. Performance test based on long-run variances and covariances.

	Canada	France	Germany	Italy	Japan	UK	USA
T	.6214 (.7413)	.4478 (.6058)	.6355 (.6972)	1.4320* (.7915)	0.5729 (.8599)	.4739 (.7066)	.4661 (.7614)
p	.2009	.2299	.1810	.0352	.2526	.2512	.2702

Table 4. Jobson-Korkie performance test.

Table 4 once again contains the values of the test statistic T_{j-1} and the corresponding standard errors, but now calculated on the basis of sample variances and covariances according to the Jobson-Korkie test. The star indicates that the corresponding Sharpe ratio difference is significantly nonnegative on a 5% level. We conclude that the MSCI index 'Italy' appears to be significantly worse than the EWP of all MSCI indices. However, this result is based on the wrong assumption of normality and serial independence of monthly asset returns. All in all it seems to be very difficult to validate portfolio strategies only by historical data. Instead, the strategies should be extensively validated by the application of Monte Carlo methods (see, e.g., Memmel, 2004, Section 5.2) rather than historical simulation. We can use the presented hypothesis test to judge whether a suggested portfolio strategy dominates some other strategies significantly, as already mentioned in Section 1.2.

3. Conclusion

In many practical situations we are searching for the best alternative among several candidates. If our decision is based on historical or simulated data there is some sort of selection bias and it is not evident if our choice is significantly optimal over all given alternatives. This problem frequently occurs in statistical inference or decisions under uncertainty such as portfolio optimization. Of course, such kind of decisions have to be reliable and thus we need a strong statistical fundament to justify our choice. In the present work a large sample test for the best alternative has been derived in a rather general setting and it has been demonstrated by an application to financial data. It was shown that the traditional Jobson-Korkie test can be generalized to ergodic stationary stochastic processes satisfying Gordin's condition. The presented hypothesis test accounts for conditional heteroscedasticity and non-normality of asset returns. We find that ignoring these kinds of stylized facts of empirical finance can lead to false rejections of the null hypothesis and misleading decisions.

Acknowledgements

I would like to thank Christoph Memmel for interesting discussions about portfolio optimization, estimation risk, and performance measurement. Many thanks belong also to Friedrich Schmid for his important suggestions.

References

Z.D. Bai (1999), 'Methodologies in spectral analysis of large dimensional random matrices, a review', *Statistica Sinica* **9**, pp. 611–677.

P.J. Brockwell and R.A. Davis (1991), *Time Series: Theory and Methods*, Springer, second edition.

J.Y. Campbell, A.W. Lo, and A.C. MacKinlay (1997), *The Econometrics of Financial Markets*, Princeton University Press.

J. Davidson (1994), *Stochastic Limit Theory*, Oxford University Press.

G.K. Eagleson (1975), 'On Gordin's Central limit theorem for stationary processes', *Journal of Applied Probability* **12**, pp. 176–179.

J.D. Hamilton (1994), *Time Series Analysis*, Princeton University Press.

F. Hayashi (2000), *Econometrics*, Princeton University Press.

J.D. Jobson and B. Korkie (1981), 'Performance hypothesis testing with the Sharpe and Treynor measures', *Journal of Finance* **36**, pp. 889–908.

P. Jorion (1985), 'International portfolio diversification with estimation risk', *The Journal of Business* **58**, pp. 259–278.

C. Memmel (2003), 'Performance hypothesis testing with the Sharpe ratio', *Finance Letters* **1**, pp. 21–23.

C. Memmel (2004), *Schätzrisiken in der Portfoliotheorie*, Ph.D. thesis, University of Cologne, Department of Economic and Social Statistics, Germany.

R.J. Muirhead (1982), *Aspects of Multivariate Statistical Theory*, John Wiley.

M. Ogaki, K. Jang, HS. Lim, et al. (2007), *Structural Macroeconometrics*, Book yet unpublished, Department of Social Studies Education, College of Education, Inha University, Korea.

F. Schmid and R. Schmidt (2007), 'Statistical inference for Sharpe's ratio', Preprint, University of Cologne, Department of Economic and Social Statistics, Germany.

University of Cologne, Statistics and Econometrics, Meister-Ekkehart-Str. 9, 50937 Cologne, Germany. Email: frahm@statistik.uni-koeln.de, phone: +49 221 470-4267.

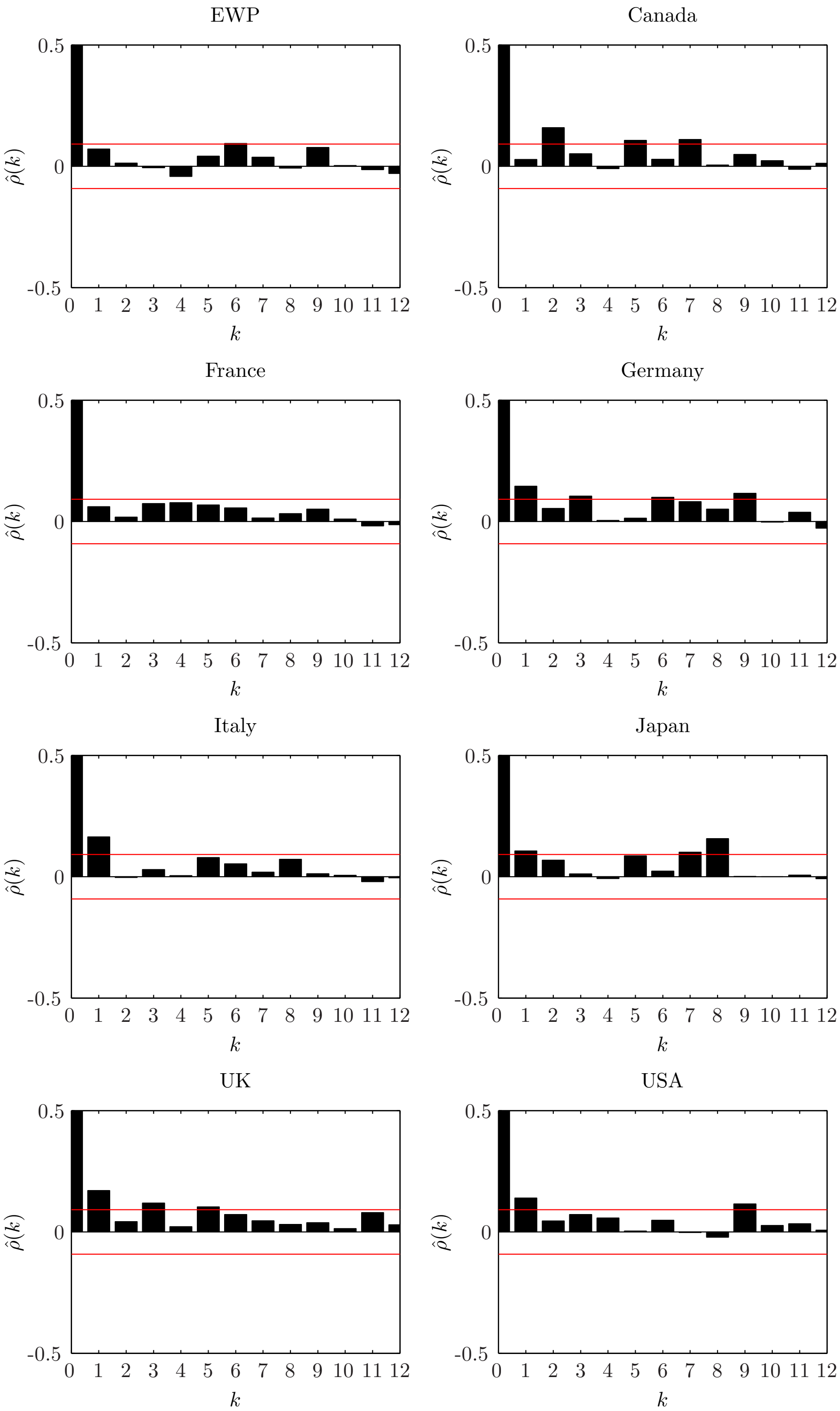

FIGURE 1. Correlograms for the squared centered excess returns of the G7 MSCI indices and the EWP. The critical thresholds for the null hypothesis $H_0\colon \rho(k) = 0$ $(k \neq 0)$ on the 5% level are indicated by the horizontal lines.