Journal of Empirical Finance 16 (2009) 816–829

Contents lists available at ScienceDirect

Journal of Empirical Finance

journal homepage: www.elsevier.com/locate/jempfin

Exact distribution-free tests of mean-variance efficiency

Sermin Gungor, Richard Luger *

Department of Economics, Emory University, Atlanta, GA 30322-2240, USA

ARTICLE INFO

Article history:
Received 6 November 2008
Received in revised form 28 April 2009
Accepted 11 June 2009
Available online 21 June 2009

JEL classification:
C12
C14
C22
G11
G12

Keywords:
CAPM
Conditional heteroskedasticity
Non-parametric tests
Robust inference

ABSTRACT

This paper develops exact distribution-free tests of unconditional mean-variance efficiency. These new tests allow for unknown forms of non-normalities, conditional heteroskedasticity, and other non-linear temporal dependencies among the absolute values of the error terms in the asset pricing model. Exactness here rests on the assumption that the joint temporal error density is symmetric around zero. This still leaves open the possibility of return distribution asymmetry via coskewness with the benchmark portfolio. A simulation study shows that the new tests have very good power relative to that of many commonly used tests. The inference procedures developed are further illustrated by tests of the mean-variance efficiency of a market index using a 42-year sample of monthly returns on ten U.S. equity portfolios.

1. Introduction

The celebrated capital asset pricing model (CAPM) of Sharpe (1964) and Lintner (1965) extends the notion of a mean-variance efficient portfolio to the portfolio of all invested wealth—the market portfolio. A given portfolio is mean-variance efficient if it has the smallest possible variance of return given its expected return, or, more appropriately, if it has the largest expected return given its variance. This theory implies that expected excess returns on assets are linearly related to the slope, or beta, of their regression on the expected excess return of the benchmark portfolio. Here excess returns are those in excess of the riskless rate of return. Under mean-variance efficiency, the risk premium of an asset is a linear function of the asset's beta.

Empirical tests of the mean-variance efficiency hypothesis are usually conducted within the context of a multivariate linear regression. The application of tests based on asymptotic theory can lead to misleading conclusions as the approximation to the finite-sample distribution of test statistics can be quite poor, especially as the number of equations included in the system increases; see Shanken (1996), Campbell et al. (1997), Dufour and Khalaf (2002), and the numerical evidence presented here. The findings show that many standard parametric tests are unreliable, rejecting the null hypothesis of mean-variance efficiency far too often.

Gibbons et al. (1989) (GRS) propose a truly finite-sample test. The exact distribution theory for their multivariate *F*-test rests on the assumption that the regression error terms are independent over time and jointly normally distributed each period, conditional on the returns of the benchmark portfolio under test. These assumptions are at odds with some well-known facts about financial asset returns. Indeed, it has long been recognized that financial returns depart from Gaussian conditions; see Fama (1965), Blattberg and Gonedes (1974), and Hsu (1982). In particular, the distribution of asset returns appears to have fatter tails than those of a normal distribution.

* Corresponding author. Tel.: +1 404 727 0328; fax: +1 404 727 4639.
E-mail address: rluger@emory.edu (R. Luger).

0927-5398/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jempfin.2009.06.003

Beaulieu et al. (2007) (BDK) propose an exact randomized likelihood-based test procedure that relaxes normality. Their framework assumes that the error distribution is known, or at least specified up to some unknown nuisance parameters. If normality is assumed, the BDK test becomes the simulation-based equivalent of the GRS test. More generally, the BDK test procedure can be thought of as an exact parametric bootstrap. When nuisance parameters are present, the computational cost of the BDK test procedure grows with the number of nuisance parameters since it then involves finding maximal p-values over a confidence region for the intervening nuisance parameters. The first-step confidence region is established by inverting a simulation-based goodness-of-fit test (of the assumed distribution), which involves a grid search over the nuisance parameter space.

Without such parametric assumptions it would seem difficult to derive an exact finite-sample distribution theory. Despite this apparent difficulty, we propose in this paper new non-randomized tests that are exact in finite samples without any parametric assumptions about the distribution of the error terms in the simple multivariate linear regression model. The methods exploit results derived by Luger (2003) in the context of testing for a random walk. Here we propose three testing approaches for joint inference on several parameters that differ mainly by what is assumed about the covariance of the errors across equations.

The first approach is an induced test procedure that allows for arbitrary covariances in the cross-section of error terms. The price to pay for this extra flexibility is that those tests can be conservative and lead to power losses if the number of test assets is large. The second approach assumes that the errors are independent across equations, conditional on the returns of the benchmark portfolio. This delivers tests with the correct size and more power than the induced tests. The third approach is based on a simple linear combination of the test assets (or portfolios of test assets) and, like the second approach, provides a test procedure with the correct size no matter the number of included assets. These single-portfolio tests allow some forms of covariation in the cross-section of error terms. The number of assets in the cross-section may even exceed the number of time-series observations, making these tests particularly attractive when testing mean-variance efficiency with many test assets or when the portfolios have relatively short histories. This stands in contrast to extant approaches based on estimates of the covariance matrix of the regression errors. In order to avoid singularities, those approaches require the size of the cross-section be less than that of the time series. The distribution-free approaches to inference developed here do not require the error covariance matrix.

The proposed distribution-free (or non-parametric) tests have several appealing features, since they are built on the mere assumption that the joint temporal error density is symmetric around zero. This means that no restrictions are placed on the degree of non-normality or the degree of heterogeneity across marginal distributions. In fact, the existence of moments need not be assumed for the validity of the new tests. It is important to note that this framework still leaves open the possibility of asymmetries in the distribution of test asset returns via coskewness with the benchmark portfolio.

Asset returns typically display clear patterns of volatility clustering for which generalized autoregressive conditional heteroskedasticity (GARCH) models are often used; see Bollerslev (1986). The tests proposed here allow not only for non-normalities, but also for *unknown* forms of conditional heteroskedasticity and other intertemporal dependencies among the absolute values of the error terms in the asset pricing model. Such forms of intertemporal dependencies invalidate the exact statistical theory underlying the parametric GRS and BDK test procedures. Since it is well known that asset returns depart from homogeneous conditions, the new tests of the mean-variance efficiency hypothesis developed here offer a valid and useful distribution-free testing alternative to potentially misleading parametric procedures.

Section 2 of the paper presents the framework, the hypothesis of interest, and the extant GRS and BDK test procedures that are exact under parametric distributional assumptions. Section 3 describes the building blocks of our three approaches to distribution-free inference. Section 4 is divided into three subsections, each one describing a proposed test procedure. Section 5 begins by presenting some additional (asymptotic) extant tests of mean-variance efficiency and then presents the results of some simulation examples to illustrate the behavior of the proposed tests relative to the commonly used procedures. Section 6 provides an empirical illustration of the new tests in the context of the Sharpe–Lintner version of the CAPM. Section 7 concludes.

2. Exact parametric tests

A benchmark portfolio with excess returns r_p is said to be mean-variance efficient with respect to a given set of N test assets with excess returns r_i, $i=1,...,N$, if it is not possible to form another portfolio of those N assets and the benchmark portfolio with the same expected return as r_p but a lower variance, or equivalently, with the same variance but a higher expected return. More formally, portfolio p is mean-variance efficient if the following first-order condition is satisfied for the N test assets:

$$E[r_{it}] = \beta_i E[r_{pt}], \quad i = 1,...,N, \tag{1}$$

where r_{it} and r_{pt} are the time-t returns on asset i and portfolio p, respectively, in excess of the riskless rate of return. The term β_i captures the degree of association between the expected excess return on the individual asset i and the expected excess return for portfolio p. Accordingly, assets with higher betas should offer in equilibrium higher expected returns. Consider the excess-return system of equations

$$r_{it} = a_i + \beta_i r_{pt} + \varepsilon_{it}, \quad t = 1,...,T, \quad i = 1,...,N, \tag{2}$$

where ε_{it} is a random error term for asset i in period t with the property that $E[\varepsilon_{it}]=0$. The specification in Eq. (2) is a seemingly unrelated equations model. The mean-variance efficiency condition in Eq. (1) can then be assessed by testing

$$H_0 : a_i = 0, \quad i = 1,...,N, \tag{3}$$

in the context of Eq. (2). This null hypothesis follows from a comparison of the unconditional expectation of Eq. (2) to the mean-variance efficiency condition in Eq. (1). If H_0 does not hold, it would be possible to obtain a higher expected return with no higher risk, contradicting the hypothesis that portfolio p is mean-variance efficient.

The exact Wald test of H_0 in Eq. (3) proposed by GRS assumes that the error terms $(\varepsilon_{1t},..., \varepsilon_{Nt})$ are jointly normally distributed around zero each period, conditional on the excess returns $(r_{p1},...,r_{pT})$; i.e., the regressor variable is assumed to be strictly exogenous. Further, the error terms are assumed independent over time. Under normality, the methods of maximum likelihood and ordinary least squares yield the same estimates of the a_i's and the β_i's in Eq. (2). Collect those estimates in $\hat{a}=(\hat{a}_1,..., \hat{a}_N)'$ and $\hat{\beta}=(\hat{\beta}_1,..., \hat{\beta}_N)'$, and let $\hat{\Sigma}$ denote the unconstrained maximum-likelihood estimate of Σ—the $N\times N$ error covariance matrix. The GRS test statistic is

$$J_1 = \frac{(T-N-1)}{N}\left[1+\frac{\hat{\mu}_p^2}{\hat{\sigma}_p^2}\right]^{-1} \hat{a}'\hat{\Sigma}^{-1}\hat{a} \ , \tag{4}$$

where $\hat{\mu}_p$ is the sample mean of r_{pt} and $\hat{\sigma}_p^2$ is the maximum-likelihood estimate of the variance of r_{pt}. Under the null hypothesis H_0, the statistic J_1 follows an exact central F distribution with N degrees of freedom in the numerator and $(T-N-1)$ degrees of freedom in the denominator.

BDK also derive exact tests of mean-variance efficiency. In the context of Eq. (2), they assume that the errors have the following distribution structure:

$$(\varepsilon_{1t},...,\varepsilon_{Nt})' = \Lambda U_t, \quad t=1,...,T, \tag{5}$$

where Λ is an unknown, non-singular matrix and the distribution of the U_t's is either completely known or specified up to a finite number of nuisance parameters, ν. BDK propose Monte Carlo (MC) tests based on the likelihood ratio (LR) statistic

$$J_2 = T(\log|\hat{\Sigma}_0| - \log|\hat{\Sigma}|), \tag{6}$$

where $\hat{\Sigma}_0$ is the constrained estimator of Σ. The key property for the MC tests is that Eq. (6) is invariant with respect to Λ. When U_t is Gaussian, the BDK test is the simulation-based equivalent of the GRS test since J_1 and J_2 are related via the monotonic transformation $J_1 = \frac{T-N-1}{N}(\exp[J_2/T]-1)$; see Equation (5.3.40) in Campbell et al. (1997). More generally, size-correct MC tests of H_0 can be performed for any assumed distribution of the U_t's that does not involve nuisance parameters. BDK also propose a maximized MC (MMC) procedure for situations with an incompletely specified error distribution, such as the multivariate Student-t distribution with unknown degrees of freedom. The MMC procedure consists of two steps. First, an exact confidence set for the nuisance parameters is constructed by inverting an MC goodness-of-fit test of the hypothesis in Eq. (5) with ν specified. Inverting that test involves assembling by grid search the values of ν that are not rejected at a specific level of significance, say α_1. In the second step, the p-value of the MC LR test is maximized over the first-step confidence set. If the resulting MMC p-value is compared to a cutoff level α_2, then the overall level of this (computationally intensive) procedure is $\alpha=\alpha_1+\alpha_2$; see BDK for details and Dufour and Khalaf (2001) for a more general discussion about the technique of MC testing. Like any MC test procedure, the BDK approach is exact only if the maintained distribution structure actually holds.

3. Building blocks

Within the context of model (2), we develop distribution-free tests of the mean-variance efficiency null hypothesis in Eq. (3). Elliptically symmetric distributions of returns are very attractive in this context because they guarantee that mean-variance analysis is fully compatible with expected utility maximization; see Chamberlain (1983), Owen and Rabinovitch (1983), and Berk (1997). If a random vector of returns follows an elliptically symmetric distribution, then the marginal distribution of any component of that vector is also elliptically symmetric; see Ingersoll (1987, p. 104). Consistent with that fact, the proposed tests are based on an assumption of symmetry for the marginal error distributions and proceed by taking long differences between return observations separated by $m=T/2$ periods. Here we assume that T is even so that the midpoint m is an integer. The approach is based on Luger (2003) and makes the following assumptions about the error terms for test asset i and the excess returns of portfolio p:

Assumption 1. The density of $(\varepsilon_{i1},..., \varepsilon_{iT})$ is symmetric around zero, conditional on $R_p=(r_{p1},..., r_{pT})'$. Further, $\Pr[r_{pt}=0]=0$ for $t=1,..., T$.

Assumption 2. The random variables ε_{it}, $t=1,..., T$, are continuous, conditional on R_p.

The approach proposed here is based on distribution-free tests about the medians of ε_i, $i=1,..., N$. Under Assumption 1, these become equivalent to the usual tests (e.g. the GRS test) that maintain $E[\varepsilon_{it}|R_p]=0$. Note that the distribution of r_{pt} may be skewed thereby inducing asymmetry in the unconditional distribution of r_{it}. In that case, a given portfolio can still be mean-variance efficient provided that investors' expected utility depends only on the mean and variance of a portfolio's return.

What distinguishes the present framework from Luger (2003) is that here the multivariate symmetry assumption is made conditional on $R_p = (r_{p1},..., r_{pT})'$. It should be noted that this conditioning on R_p is entirely consistent with arbitrage pricing theory. It is also a common approach used by GRS and BDK, for instance, to obtain a finite-sample distribution theory. The distributional assumptions made by those authors, however, are far more restrictive than Assumptions 1 and 2. GRS require independent and jointly normal errors, while BDK relax the normality assumption but nevertheless require that the error distribution structure be specified up to a finite number of nuisance parameters. Here the conditional distribution of $(\varepsilon_{i1},..., \varepsilon_{iT})$ given R_p is completely unspecified. This means that no restrictions are placed on the degree of non-normality or the degree of heterogeneity across marginal distributions. It should be noted that several popular models of time-varying conditional variance, such as GARCH-type or stochastic volatility models with continuous and symmetric innovations, satisfy Assumptions 1 and 2. In fact, those assumptions allow for the presence of *unknown* forms of conditional heteroskedasticity (or any other form of non-linear dependence among the $|\varepsilon_{it}|$'s). Furthermore, the conditional variances need not be finite nor even follow a stationary process, and moments need not exist.

Define the scaled returns r_{it}/r_{pt}, for $i = 1,..., N$ and $t = 1,..., T$. Under the statistical specification in Eq. (2), these scaled returns can be expressed as

$$\frac{r_{it}}{r_{pt}} = \frac{a_i}{r_{pt}} + \beta_i + \frac{\varepsilon_{it}}{r_{pt}},$$

thereby relegating the nuisance parameter β_i to the role of intercept. Following Theil (1971, p. 616), that parameter can be eliminated from the inference problem via the long differences

$$z_{it} = d_t^m \times x_{pt}, \tag{7}$$

for $t = 1,..., m$, where

$$d_t^m = \left(\frac{r_{i,t+m}}{r_{p,t+m}} - \frac{r_{it}}{r_{pt}}\right) \text{ and } x_{pt} = \frac{(r_{pt} - r_{p,t+m})}{r_{pt} r_{p,t+m}}.$$

The quantity in Eq. (7) is the building block of the distribution-free methods proposed here. An important aspect of the sequence of long differences d_t^m, $t = 1,..., m$, is that it contains no overlaps. The methodology of Luger (2003) yields exact tests when built on d_t^m. The multiplication of d_t^m by x_{pt} seen in Eq. (7) is done in order to improve power. To see why this is, note that

$$d_t^m = a_i x_{pt} + \left(\frac{\varepsilon_{i,t+m}}{r_{p,t+m}} - \frac{\varepsilon_{it}}{r_{pt}}\right). \tag{8}$$

The proof of Proposition 1 below shows that $(\varepsilon_{i,t+m}/r_{p,t+m} - \varepsilon_{it}/r_{pt})$ has a symmetric distribution. Suppose that $x_{pt} = (1/r_{p,t+m} - 1/r_{pt})$ is also symmetrically distributed, independently of the ε_{it}'s. Then the quantity in Eq. (8) would have median zero no matter the value of a_i, so that any sign statistic based on that quantity would have nothing but trivial power. Multiplying d_t^m with x_{pt} to get z_{it} shifts its median to the right or left, depending on the sign and magnitude of a_i, and hence delivers sign tests with power to detect $a_i \neq 0$. While such a modification results in only approximately exact tests in Luger's (2003) random walk context, here the tests are based on truly exact finite-sample pivots owing to the fact that we can condition on R_p.

Define a sign function as $s[z] = 1$ if $z > 0$, and $s[z] = 0$ if $z \leq 0$. The following two propositions are adapted from Luger (2003) and proofs are provided in Appendix A for the sake of self-containment. Note that when $a_i = 0$, the statistic z_{it} in Eq. (7) is a function only of $(\varepsilon_{i,t+m}/r_{p,t+m} - \varepsilon_{it}/r_{pt})x_{pt}$.

Proposition 1. *Let $\varepsilon_{i1},..., \varepsilon_{iT}$ be a sequence of random variables that satisfies Assumptions 1 and 2. Then, the random variable $s[(\varepsilon_{i,t+m}/r_{p,t+m} - \varepsilon_{it}/r_{pt})x_{pt}]$ is distributed like a Bernoulli variable B_t, such that $\Pr[B_t = 1] = \Pr[B_t = 0] = 1/2$ for $t = 1,2,..., m$.*

This result shows that the random variables $s[z_{it}]$, $t = 1,...,m$, are identically distributed under the null hypothesis. The next proposition shows that those random variables are also mutually independent, thereby paving the way for test procedures based on a class of linear signed rank statistics defined by

$$SR_i = \sum_{t=1}^{m} s[z_{it}]\varphi[Rank(|z_{it}|)], \tag{9}$$

where $Rank(|z_{it}|)$ is the rank of $|z_{it}|$ when $|z_{i1}|,..., |z_{im}|$ are placed in ascending order of magnitude, and $0 \leq \varphi[1] \leq ... \leq \varphi[m]$ is a set of scores with $\varphi[m] > 0$. In the following, the symbol $\overset{d}{=}$ stands for the equality in distribution.

Proposition 2. *Let $\varepsilon_{i1},\ldots,\varepsilon_{iT}$ be a sequence of random variables that satisfies Assumptions 1 and 2. Then, the null distribution of any linear signed rank statistic defined by Eq. (9) has the property that*

$$SR_i \overset{d}{=} \sum_{j=1}^{m} B_j \varphi[j],$$

where $B_1,\ldots, B_m$ are mutually independent uniform Bernoulli variables on {0,1}.

The choice of score function yields different linear signed rank statistics. The two statistics of greatest interest are the sign statistic, obtained from the score function $\varphi[j]=1$:

$$S_i = \sum_{t=1}^{m} s[z_{it}],$$

and the Wilcoxon signed rank statistic

$$W_i = \sum_{t=1}^{m} s[z_{it}] Rank(|z_{it}|),$$

obtained with $\varphi[j]=j$. The sign statistic follows a binomial distribution with number of trials m and probability of success 1/2. The distribution of the Wilcoxon statistic has been tabulated for various values of m; see Table A.4 in Hollander and Wolfe (1973) for $m\leq 15$. For larger values, the standard normal distribution provides a very good approximation to the standardized variate. The same is true of the standardized sign statistic. The basic result (found in Randles and Wolfe 1979, Section 10.2) is that the standardized sign and Wilcoxon signed rank statistics

$$S_i^* = \frac{S_i - m/2}{\sqrt{m/4}} \text{ and } W_i^* = \frac{W_i - m(m+1)/4}{\sqrt{m(m+1)(2m+1)/24}}, \tag{10}$$

where null means and variances are used, have limiting (as $m\to\infty$) standard normal distributions. It should be noted that the asymptotic approximation works extremely well even for small values of m since the Binomial and Wilcoxon distributions are close to normal.

In principle, the approach adopted here could be extended to allow for additional covariates. For example, suppose returns were represented by

$$r_{it} = a_i + \beta_i r_{pt} + \gamma_i r_{ft} + \varepsilon_{it}, \quad t = 1,\ldots,T, \quad i = 1,\ldots,N,$$

where r_{ft} is a second common factor. In that case, the basic building of the distribution-free tests would become

$$z_{it} = \left(\frac{d_{t+m/2}^m}{x_{f,t+m/2}} - \frac{d_t^m}{x_{ft}}\right)\left(\frac{x_{p,t+m/2}}{x_{f,t+m/2}} - \frac{x_{pt}}{x_{ft}}\right),$$

where $x_{ft} = (r_{pt}r_{f,t+m} - r_{ft}r_{pt+m})/r_{pt}r_{p,t+m}$ for $t=1,\ldots, m/2$.

4. Exact distribution-free tests

4.1. Sup-type tests

A test of H_0 with a single test asset can be performed simply by comparing either of the test statistics in Eq. (10) with the appropriate critical values from the standard normal distribution. In order to test the null hypothesis with several test assets indexed by $i=1,\ldots, N$, consider the test statistics

$$SB = \max_{1\leq i\leq N} |S_i^*| \text{ and } WB = \max_{1\leq i\leq N} |W_i^*|. \tag{11}$$

Note that the maximal statistic corresponds to the one with the smallest p-value, since the individual test statistics are identically distributed. These test statistics are suggested by the fact that H_0 in Eq. (3) is the intersection of the N subhypotheses $H_{0i}: a_i=0$, $i=1,\ldots, N$. The decision rule is then built from the logical equivalence that H_0 is false if any of its component subhypotheses is false; i.e., reject H_0 if any one of the separate tests, say $S_1^*,\ldots, S_N^*$, rejects it. Such tests are called *induced* tests of H_0;

see Savin (1984). The results of Sidak (1967) show that no matter the covariance structure among the individual test statistics, the asymptotic marginal null distributions of *SB* and *WB* defined in Eq. (11) satisfy the inequalities

$$Pr[SB \le \omega_{\alpha^*/2}] \ge (1-\alpha) \text{ and } Pr[WB \le \omega_{\alpha^*/2}] \ge (1-\alpha), \tag{12}$$

where $\omega_{\alpha^*/2}$ is the upper $\alpha^*/2$ critical point of the standard normal distribution and $\alpha^* = 1-(1-\alpha)^{1/N}$. This inequality gives a slight improvement over the Bonferroni inequality. The inequalities in Eq. (12) imply that the asymptotic *level* of the induced test of H_0 that compares either *SB* or *WB* to $\omega_{\alpha^*/2}$ is equal to α. So if the ordinary two-sided *p*-value of *SB* or *WB* is, say $p\nu$, then the multiplicity-adjusted two-sided *p*-value is calculated from the equation $p\nu^* = 1-(1-p\nu)^N$. See Chow and Denning (1993) for an application of Sidak's results in a different context.

4.2. Sum-type tests

The sup-type tests based on the bounds in Eq. (12) allow for an arbitrary covariance structure in the cross-section of error terms. The cost of this extra flexibility is that those tests tend to be conservative and this behavior under the null effectively translates into relative power losses under the alternative hypothesis; see Section 4 for some simulation evidence.

Tests with the correct *size* that deliver more power can be obtained under the assumption that the error terms $(\varepsilon_{1t},..., \varepsilon_{Nt})$ are independent, conditional on $R_p = (r_{p1},..., r_{pT})'$. Note that the error terms need not be unconditionally independent. With that further assumption, the exact distributions of the sum-type statistics

$$SS = \sum_{i=1}^{N} S_i^{*2} \text{ and } WS = \sum_{i=1}^{N} W_i^{*2} \tag{13}$$

can be obtained (e.g. by numerical or simulation methods). A far more practical way to conduct inference is to note that since the S_i^*'s and W_i^*'s tend to be normal as *m* increases, the null distributions of *SS* and *WS* in Eq. (13) will tend to be chi-squared with *N* degrees of freedom. The results in Section 4 show that this approximation works extremely well.

4.3. Single-portfolio tests

As GRS remark, the hypothesis that $a_i = 0$, for all *i*, implies that all linear combinations of the a_i's equal zero. This means that if some linear combination of the *N* test assets (or portfolios of test assets) has a non-zero intercept, then the null hypothesis is false; i.e., portfolio *p* is not mean-variance efficient. Following that idea, consider aggregating the *N* equations that comprise the system in Eq. (2) to find the single-portfolio representation

$$\bar{r}_t = A + Br_{pt} + e_t, \quad t = 1,..., T, \tag{14}$$

where $\bar{r}_t = \sum_{i=1}^{N} r_{it}$, $A = \sum_{i=1}^{N} a_i$, $B = \sum_{i=1}^{N} \beta_i$, and $e_t = \sum_{i=1}^{N} \varepsilon_{it}$. It is also possible to consider weighted sums. The weights, however, would need to be chosen on the basis of prior (non-sample) information to avoid introducing data-snooping size distortions; see Lo and MacKinlay (1990).

The construction of a test of H_0 in Eq. (3) based on *A* is reminiscent of the test in Bossaerts and Hillion (1995) based on $\hat{A} = \sum_{i=1}^{N} \hat{a}_i$. The same distribution-free approach as above can be used to detect whether the intercept *A* in Eq. (14) is different from zero under the following high-level assumptions:

Assumption 3. The density of $(e_1,..., e_T)$ is symmetric around zero, conditional on $R_p = (r_{p1},..., r_{pT})'$. Further, $\Pr[r_{pt} = 0] = 0$ for $t = 1,..., T$.

Assumption 4. The random variables e_t, $t = 1,..., T$, are continuous, conditional on R_p.

These assumptions are the natural extensions of Assumptions 1 and 2 to the model in Eq. (14). As before, no restrictions are placed on the shape or the location of the distribution of r_{pt}. It is easy to see that if Assumption 1 holds and the errors are either cross-sectionally independent or at least exchangeable as Bossaerts and Hillion (1995) assume, then Assumption 3 is satisfied. A sufficient condition for Assumption 3 is that the joint cross-sectional density of error terms is reflective symmetric so that $(\varepsilon_{1t},..., \varepsilon_{Nt})$ has the same distribution as $(-\varepsilon_{1t},..., -\varepsilon_{Nt})$. In that case, we have that $\sum_{i=1}^{N} \varepsilon_{it}$ has the same distribution as $-\sum_{i=1}^{N} \varepsilon_{it}$; i.e., e_t is symmetrically distributed around zero. So although not completely unrestricted, reflective symmetry allows some forms of covariation in the cross-section of error terms.

Our distribution-free approach yields the following versions of the sign and Wilcoxon test statistics:

$$S = \sum_{t=1}^{m} s[Z_t] \text{ and } W = \sum_{t=1}^{m} s[Z_t] Rank(|Z_t|), \tag{15}$$

where $Rank(|Z_t|)$ is the rank of $|Z_t|$ when $|Z_1|,...,|Z_m|$ are placed in ascending order of magnitude. Here Z_t is defined as

$$Z_t = \left(\frac{\bar{r}_{t+m}}{r_{p,t+m}} - \frac{\bar{r}_t}{r_{pt}}\right)\left(\frac{r_{pt} - r_{p,t+m}}{r_{pt} r_{p,t+m}}\right),$$

for $t = 1,...,m$. Under the null hypothesis, the sign and Wilcoxon statistics in Eq. (15) based on Z_t follow exactly the binomial distribution (with parameters m and 1/2) and the distribution of the Wilcoxon variate, respectively. As before, the standardized versions

$$S^* = \frac{S - m/2}{\sqrt{m/4}} \text{ and } W^* = \frac{W - m(m+1)/4}{\sqrt{m(m+1)(2m+1)/24}} \tag{16}$$

have limiting standard normal distributions under H_0. Of course, the signed rank tests built on Z_t have no power to detect mean-variance inefficiencies of portfolio p in directions away from the null where $\sum_{i=1}^{N} a_i = 0$. This possibility, however, can easily be checked by examining the estimates $\hat{a}_1,..., \hat{a}_N$; see Section 5 for an empirical illustration with size portfolios.

The possibility of losing power depending on whether the (weighted) a_i's tend to cancel out is a concern for any approach that uses linear combinations of assets. For example, parametric tests of the CAPM are usually applied to portfolio groupings of stocks in order to have N much less than T. As Shanken (1996) notes, this has the potential effect of reducing the residual variances and increasing the precision with which the a_i's are estimated. On the other hand, as Roll (1979) points out, individual stock expected return deviations under the alternative can cancel out in portfolios, which would reduce power. So the expected power of tests to detect mean-variance inefficiencies necessarily depends on how the portfolios are formed; see Sentana (2008) for more on the effects of portfolio composition on test power.

5. Simulation results

This section reports the results of some simulation experiments to contrast the behavior of the proposed distribution-free tests with several standard test procedures. The first of these benchmarks is the GRS test in Eq. (4). The second one, denoted J_2^{MC}, is an MC test based on the LR statistic in Eq. (6), following the BDK approach. The other benchmarks for comparison purposes are the usual LR test, an adjusted LR test, and a test based on the Generalized Method of Moments (GMM). The latter is a particularly important benchmark here, since in principle it is "robust" to non-normality and heteroskedasticity of returns.

The LR test, J_2, is applied following the usual practice which is to base it on asymptotic theory. The large-sample result is that the null distribution of J_2 tends to that of a chi-square variate with N degrees of freedom, χ^2_N. Jobson and Korkie (1982) suggest an adjustment to J_2 in order to improve its finite-sample size properties when compared against the χ^2_N distribution. The adjusted statistic is

$$J_3 = \frac{T - (N/2) - 2}{T} J_2, \tag{17}$$

which, under H_0, also follows an asymptotic χ^2_N distribution.

MacKinlay and Richardson (1991) develop tests of mean-variance efficiency using a GMM framework. Define r_t as the $N \times 1$ vector of excess returns for the N test assets observed at time t. The moments of interest are those of the $2N \times 1$ vector

$$f_t(\theta) = h_t \otimes \varepsilon_t(\theta), \tag{18}$$

where $h_t = (1, r_{pt})'$ and $\varepsilon_t(\theta) = r_t - a - \beta r_{pt}$. Here $\theta = (a', \beta')'$, $a = (a_1,..., a_N)'$, and $\beta = (\beta_1,..., \beta_N)'$. The symbol $\otimes$ refers to the Kronecker product. The model specification in Eq. (2) implies the moment conditions $E(f_t(\theta_0)) = 0$, where θ_0 is the true parameter vector. The system in Eq. (18) is exactly identified which implies that the GMM procedure yields the same estimates of θ as does OLS applied equation by equation, which in turn are equivalent to the maximum likelihood estimates in Eq. (2). The GMM-based Wald test statistic is

$$J_4 = T \hat{a}' \left[R \left(\hat{D}' \hat{S}^{-1} \hat{D} \right)^{-1} R' \right]^{-1} \hat{a} \ , \tag{19}$$

where $R = (1,0) \otimes I_N$, with I_N as the $N \times N$ identity matrix. Here $T^{-1}(\hat{D}' \hat{S}^{-1} \hat{D})^{-1}$ is a consistent estimator of the covariance matrix of the GMM estimator $\hat{\theta}$. The component matrix $\hat{D}$ is computed as

$$\hat{D} = -\begin{bmatrix} 1 & \hat{\mu}_p \\ \hat{\mu}_p & (\hat{\sigma}_p^2 + \hat{\mu}_p^2) \end{bmatrix} \otimes I_N$$

and $\hat{S}$, which is a consistent estimator of $\sum_{\ell=-\infty}^{+\infty} E[f_t(\theta_0)f_{t-\ell}(\theta_0)']$, can be computed with the Newey and West (1987) procedure. Under the null hypothesis H_0, the statistic J_4 follows an asymptotic χ^2_N distribution. Following MacKinlay and Richardson (1991), the J_4 statistic in Eq. (19) is scaled by $(T-N-1)/T$, which was suggested to improve the finite-sample null behavior of the GMM Wald test.

We also consider the parametric J tests applied to the single-portfolio representation in Eq. (14). Those tests are denoted J_1^*, J_2^{*MC}, J_2^*, J_3^*, and J_4^* and they will allow us to see the relative power gains that result from the non-parametric approach versus those that the linear combination induces. Note that since $\log(1+x)\approx x$ for x small, we expect J_3^* to mimic J_1^*.

All the tests compared here are invariant with respect to the β_i's, so there is no need to specify their values. The MC tests are applied assuming the distribution of the error terms is completely known and the p-values are based on 100 MC draws. We consider sample sizes $T=30$, 60, 120, and 240. For the number of test assets, we consider values of $N=10$, 20, and 30. Note that J_1, J_2^{MC}, J_2, J_3, and J_4 cannot be computed when $N\geq T$ (Those cases are abbreviated n.a. in the tables). The GMM Wald test, J_4, cannot even be computed with $T=60$ and $N=30$. The reason is that the matrix $\hat{S}$ is singular whenever $2N=T$. All the other tests do not have these limitations on N relative to T. In order to have a sharp contrast between the power functions of each test, the parameters a_i are set to 0.07 under the alternative hypothesis. The parametric MC tests are exact here, but there is no such theoretical guarantee with all the other parametric tests across error specifications. So to ensure meaningful comparisons, the power results for the non-randomized parametric tests are based on size-corrected critical values. Finally, each experiment comprises 1000 replications of each return generating process.

Two specifications of model (2) are considered. The first example, though completely unrealistic, is constructed to ensure the exactness of the GRS Wald test, J_1. The error terms ε_{it} are drawn randomly from the standard normal distribution and so are portfolio-p's excess returns, r_{pt}, independently of the errors. MacKinlay and Richardson (1991) use a similar design. We also tried introducing coskewness by drawing r_{pt} from a χ^2_1 distribution. The results were essentially the same as in the symmetric case.

Table 1 reports the empirical rejection rates for this homoskedastic example. As expected, the parametric J_1 and J_2^{MC} tests, the distribution-free sum-type *SS* and *WS* tests, and the single-portfolio J_1^*, J_2^{*MC}, S^*, and W^* tests have empirical sizes close to the nominal 5% level. Note how well the use of critical values from the chi-squared distribution works for the *SS* and *WS* tests, and those of the normal distribution for the S^* and W^* tests. In general, all the single-portfolio J^* tests also have rejection rates close to the nominal level. We see though that even in this i.i.d. case, the LR J_2 test suffers serious size distortions when T is relatively small and N increases. For example, when N increases from 10 to 20 with $T=30$, the empirical size of the J_2 test increases from about 15% to 60%—twelve times the nominal level.

The bottom portion of Table 1 shows the relative power results. Note that J_1, J_2, and J_3 have identical size-corrected powers, since they are all related via monotonic transformations. The same is true of their single-portfolio counterparts. As expected, the J_2^{MC} and J_2^{*MC} tests behave like the J_1 and J_1^* tests, respectively. It is immediately clear that *SB* and *WB* are not as powerful as the other tests. They nevertheless have power as T increases. The most striking result though comes from comparing the S^* and W^*

Table 1
Empirical size and power comparisons in the seemingly unrelated equations model with homoskedastic errors.

N	T	J_1	J_2^{MC}	J_2	J_3	J_4	*SB*	*WB*	*SS*	*WS*	J_1^*	J_2^{*MC}	J_2^*	J_3^*	J_4^*	S^*	W^*
Under $H_0: a_i=0$, $i=1,\ldots,N$																	
10	30	5.0	4.7	16.5	5.0	13.8	6.6	2.6	3.9	4.0	4.8	5.5	6.4	4.7	6.9	2.5	4.2
	60	4.0	4.8	7.4	4.0	6.4	5.5	3.5	4.8	4.4	4.0	4.0	4.3	4.0	4.4	5.1	4.9
	120	4.9	4.9	7.3	4.9	6.5	6.3	5.5	5.3	5.1	3.9	4.0	4.1	3.9	4.0	4.5	5.0
	240	3.8	3.7	4.6	3.8	4.6	3.0	4.5	4.1	4.6	3.9	3.7	3.9	3.9	3.9	4.5	4.2
20	30	4.7	4.9	63.1	12.4	n.a.	2.1	1.8	3.3	3.5	4.7	4.3	5.9	4.7	6.1	4.2	4.9
	60	4.7	5.3	22.8	5.3	12.7	2.8	3.1	4.6	4.7	4.5	4.2	5.1	4.5	5.2	4.1	4.6
	120	5.1	5.1	11.5	5.2	8.8	5.1	3.7	4.2	4.6	6.0	6.0	6.3	6.0	6.3	4.7	4.5
	240	5.3	4.9	6.6	5.3	5.9	5.0	4.5	4.8	4.8	4.7	5.4	4.9	4.7	4.8	5.3	5.2
30	30	n.a.	n.a.	n.a.	n.a.	n.a.	2.8	1.2	4.6	3.1	4.4	4.5	5.7	4.4	5.7	3.5	4.1
	60	4.4	4.2	45.9	6.6	n.a.	3.1	3.0	4.7	5.0	4.8	4.7	5.5	4.8	5.6	3.8	5.4
	120	4.8	4.8	16.3	5.4	8.8	2.8	3.2	4.6	4.7	4.3	5.0	4.4	4.3	4.3	5.2	4.7
	240	4.5	4.5	8.5	4.5	6.5	3.6	4.9	5.3	5.0	5.1	5.6	5.3	5.1	5.2	5.5	5.1
Under $H_1: a_i=0.07$, $i=1,\ldots,N$																	
10	30	7.0	7.2	7.0	7.0	7.9	6.8	2.8	4.2	5.6	20.6	20.3	20.6	20.6	20.4	6.9	10.4
	60	16.3	13.7	16.3	16.3	16.7	6.6	5.1	7.6	9.3	40.5	34.7	40.5	40.5	40.0	11.2	13.7
	120	29.0	26.6	29.0	29.0	28.3	9.0	8.8	8.9	10.2	71.3	64.5	71.3	71.3	70.9	23.2	26.6
	240	63.2	58.5	63.2	63.2	63.4	12.0	13.4	16.8	20.4	93.6	92.2	93.6	93.6	93.6	45.7	49.6
20	30	7.3	6.9	7.3	7.3	n.a.	2.1	2.2	5.4	6.3	38.8	36.2	38.8	38.8	38.5	12.0	16.0
	60	16.2	15.1	16.2	16.2	16.3	3.7	4.5	7.4	8.0	66.0	63.5	66.0	66.0	65.8	21.0	27.0
	120	43.2	41.3	43.2	43.2	43.4	8.8	9.5	11.5	14.0	91.6	92.4	91.6	91.6	91.7	43.5	49.0
	240	83.1	83.2	83.1	83.1	82.8	12.7	13.6	25.6	28.5	99.9	99.9	99.9	99.9	99.9	70.0	75.5
30	30	n.a.	n.a.	n.a.	n.a.	n.a.	2.8	1.5	6.1	7.2	51.2	47.4	51.2	51.2	50.5	13.7	17.4
	60	19.5	18.3	19.5	19.5	n.a.	5.3	3.4	6.8	8.8	83.2	82.8	83.2	83.2	82.8	27.6	34.5
	120	50.0	48.0	50.0	50.0	52.0	6.0	8.9	15.0	17.8	99.1	98.8	99.1	99.1	99.1	56.9	63.3
	240	92.1	90.7	92.1	92.1	92.2	9.8	12.4	28.7	34.5	100.0	100.0	100.0	100.0	100.0	85.1	90.8

Note: Nominal level is 0.05 and entries are percentage rates. Results based on 1000 replications. The abbreviation n.a. stands for not applicable.

tests with the parametric tests. For example, the W^* test has power that is either comparable or even greater in some instances than all the parametric J tests. The power advantage that the linear combination induces in this setup is apparent when comparing any test to its single-portfolio counterpart. The power performance of the S^* and W^* tests is quite remarkable considering their non-parametric nature and the fact that only half the sample is effectively used to detect departures from the null hypothesis.

The second specification of model (2) is a more realistic description of asset returns. Following MacKinlay and Richardson (1991), we consider a case of contemporaneous conditional heteroskedasticity in which the variance of ε_{it} depends on the value of r_{pt}. In their example, the error terms are essentially governed by an equation of the form $\varepsilon_{it} = \eta_{it}\sqrt{(r_{pt}-\mu_p)^2 / \sigma_p^2}$, where η_{it} is the innovation term. As in standard GARCH models, this form does not allow an asymmetric response to shocks in r_{pt}; i.e., the conditional variance of ε_{it} is the same whether $(r_{pt}-\mu_p)$ is positive or negative. Here we consider a specification in the spirit of Nelson's (1991) exponential GARCH model that allows for asymmetric responses to shocks. The innovations η_{it} are standard normal and we let $\varepsilon_{it} = \eta_{it}\sqrt{exp(\lambda_0 + \lambda_1 r_{pt})}$. It should be noted that such a specification finds empirical support in Duffee (1995, 2001). The returns r_{pt} are as in the first example and we set $\lambda_0 = 0$, $\lambda_1 = 2$. It is easy to see that this specification generates ε_{it}'s with excess kurtosis.

Table 2 reports the results for the heteroskedastic example. From the top portion where $a_i = 0$, we see that all the non-randomized parametric J tests suffer serious size distortions, and that these worsen as N increases for any given T. When $N = 10$, $T = 60$ the parametric J tests all have empirical sizes in excess of 20%. The probability of a Type I error with J_1, J_2, and J_3 exceeds 60% when N is increased to 30. The severity of the size distortions for the GMM-based J_4 test is also quite surprising since it accommodates conditional heteroskedasticity, at least in principle. In sharp contrast, there is a closer agreement between the nominal and empirical rejection probabilities for the single-portfolio J^* tests. The MC tests do not have an over-rejection problem here, since they are applied assuming the form of heteroskedasticity is known. The distribution-free tests also behave as expected with rejection rates close to the nominal 5% level, but unlike the MC tests this is achieved without assuming anything about the heterogeneity in variances.

From the bottom portion of Table 2, we see that the distribution-free tests have very good power when compared to the parametric tests. The sum-type SS and WS tests have power which is comparable to that of the J tests, including the MC tests. It should be emphasized that the size-corrected tests are not feasible tests in practice. They are merely used here as theoretical benchmarks for the new tests. Comparing the single-portfolio tests, we see again that the S^* and W^* tests in this heteroskedastic case tend to be more powerful than the parametric tests, often by a sizeable margin. This result is in line with the discussion in Randles and Wolfe (1979, Ch. 4) about the power of the sign and Wilcoxon signed rank tests relative to the Student-t test in a location model under a variety of distributions. Here we have another example that illustrates the conventional wisdom that non-parametric tests tend to perform well in the presence of excess kurtosis—a well-known feature of asset returns.

Table 2
Empirical size and power comparisons in the seemingly unrelated equations model with contemporaneous heteroskedastic errors.

N	T	J_1	J_2^{MC}	J_2	J_3	J_4	SB	WB	SS	WS	J_1^*	J_2^{*MC}	J_2^*	J_3^*	J_4^*	S^*	W^*
Under $H_0: a_i = 0$, $i = 1,\ldots, N$																	
10	30	25.9	4.6	50.7	27.7	41.6	6.0	2.1	3.8	4.3	8.1	4.7	9.6	8.1	7.4	2.9	3.7
	60	23.0	5.0	31.4	23.2	25.4	3.8	3.4	3.5	4.2	6.6	4.1	7.4	6.6	6.4	2.7	4.2
	120	15.5	5.8	19.5	15.5	15.5	5.8	4.2	3.7	4.3	6.1	5.4	6.5	6.1	5.4	5.6	4.4
	240	9.6	4.9	10.9	9.6	9.0	3.2	3.7	4.4	4.2	4.4	5.2	4.5	4.4	4.6	4.5	4.3
20	30	45.2	4.1	96.1	62.9	n.a.	2.0	1.8	4.6	4.9	7.0	4.9	8.3	7.0	6.8	4.1	5.4
	60	45.1	4.8	66.8	46.7	55.3	3.5	3.2	4.7	5.0	6.5	4.5	7.1	6.5	5.2	2.8	4.4
	120	29.9	4.5	41.0	30.2	33.1	5.5	4.5	5.6	5.4	6.7	5.0	6.8	6.7	5.5	5.4	4.1
	240	16.9	4.8	20.3	16.9	17.1	4.7	5.0	5.1	4.2	3.8	4.2	3.9	3.8	3.9	5.6	5.4
30	30	n.a.	n.a.	n.a.	n.a.	n.a.	2.7	1.7	4.2	4.2	6.1	5.0	7.1	6.1	6.6	4.2	4.1
	60	64.3	5.0	95.0	69.9	n.a.	3.9	2.6	4.6	4.2	6.5	5.3	6.6	6.5	6.4	4.5	4.6
	120	46.7	5.5	65.3	48.2	56.1	3.5	4.4	5.3	5.6	5.5	4.4	5.7	5.5	5.3	5.4	5.5
	240	27.6	5.2	37.7	28.1	28.3	3.3	4.3	5.5	4.9	4.6	5.2	4.7	4.6	4.7	5.1	3.4
Under $H_1: a_i = 0.07$, $i = 1,\ldots, N$																	
10	30	7.2	5.6	7.2	7.2	7.2	7.1	2.3	4.6	4.9	8.0	9.9	8.0	8.0	9.5	4.9	7.8
	60	8.6	8.2	8.6	8.6	7.9	6.1	5.6	6.3	6.2	12.0	11.8	12.0	12.0	13.1	10.4	11.2
	120	9.0	10.0	9.0	9.0	9.1	9.9	7.0	7.6	7.5	17.5	18.6	17.5	17.5	19.3	17.2	20.2
	240	11.1	11.9	11.1	11.1	14.2	8.5	10.4	12.5	13.7	29.7	28.9	29.7	29.7	31.0	32.3	37.7
20	30	5.0	8.2	5.0	5.0	n.a.	2.9	2.4	6.3	5.6	12.9	13.1	12.9	12.9	13.4	8.6	11.8
	60	7.6	9.6	7.6	7.6	7.4	4.1	4.5	7.1	7.4	15.8	18.0	15.8	15.8	19.9	14.5	18.8
	120	11.7	11.8	11.7	11.7	12.1	8.7	7.9	9.6	12.3	24.5	31.5	24.5	24.5	32.1	31.0	37.6
	240	19.2	20.9	19.2	19.2	20.7	9.6	10.6	15.9	19.7	54.1	48.7	54.1	54.1	56.2	56.7	65.8
30	30	n.a.	n.a.	n.a.	n.a.	n.a.	2.9	1.0	5.2	4.9	16.2	18.5	16.2	16.2	21.6	9.3	13.7
	60	8.7	11.5	8.7	8.7	n.a.	4.4	3.0	7.9	7.3	24.8	26.4	24.8	24.8	26.0	21.5	29.8
	120	11.3	14.9	11.3	11.3	12.3	5.1	7.0	10.1	11.4	39.9	40.8	39.9	39.9	43.2	40.1	47.8
	240	25.9	27.7	25.9	25.9	32.1	9.6	12.3	21.2	25.6	62.5	62.1	62.5	62.5	64.3	71.8	81.4

Note: The MC tests, J_2^{MC} and J_2^{*MC}, are applied assuming the form of heteroskedasticity is known. Nominal level is 0.05 and entries are percentage rates. Results based on 1000 replications. The abbreviation n.a. stands for not applicable.

6. Empirical illustration

The new distribution-free inference procedures are illustrated by tests of the mean-variance efficiency of a stock market index: the value-weighted index from the University of Chicago's Center for Research in Security Prices (CRSP). The data consist of monthly returns on ten portfolios for the period covering January 1965 to December 2006. Stocks listed on the New York Stock Exchange and on the American Stock Exchange are allocated to 10 portfolios based on the market value of equity. The one-month U.S. Treasury bill return is used as the riskless rate of return. These data definitions are the same as in Campbell et al. (1997), except that we extend the sample end point.

Table 3 shows the empirical test results for the entire 42-year period, seven 5-year subperiods and a 7-year subperiod, and three 10-year subperiods and a 12-year subperiod. It is quite common in this literature to test the CAPM over subperiods out of concerns about parameter stability. Here the choice of subperiods follows that in Campbell, Lo, and MacKinlay. Columns 2–6 report the results of the parametric J tests; columns 7 and 8 report the results of the sup-type SB and WB tests; columns 9 and 10 report the results of the sum-type SS and WS tests; columns 11–15 report the results of the single-portfolio J^* tests; and columns 16 and 17 report the results of the single-portfolio S^* and W^* tests. Here the MMC test procedure of BDK is applied assuming Student-t errors with unknown degrees of freedom. In each case, we set $\alpha_1 = 0.025$ so the level of the first-step confidence interval for the degrees-of-freedom parameter is 97.5%. These confidence intervals are reported in square brackets. The reported MMC p-values should then be referred to a 2.5% cutoff if one has in mind an overall 5%-level test.

The parametric J test results tend to reject the Sharpe–Lintner CAPM, with p-values less than 6% for the entire sample period. The J^* tests reject the null hypothesis with even smaller p-values in this case. The exceptions among the parametric J tests are the MMC tests which tend to favor a non-rejection. Note that the confidence intervals associated with the MMC tests are tight at quite low values for the degrees-of-freedom parameter. This suggests the presence of high kurtosis and shows that normality is definitely rejected. In contrast, the p-values for the distribution-free tests provide very clear evidence in favor of mean-variance efficiency.

The parametric J tests and distribution-free tests results align more closely over the 5-year subperiods with both sets of tests tending to reject the null in the first three 5-year subperiods, then tending to not reject over the next four 5-year subperiods, and finally tending to reject again the null in the last 7-year subperiod. The single-portfolio J^* tests tend to be in agreement with the other tests in every one of those subperiods except in the subperiod 1/70–12/74, where the J^* tests indicate a clear non-rejection

Table 3
Mean-variance efficiency tests of the Sharpe–Lintner version of the CAPM.

Time	J_1	J_2^{MMC}	J_2	J_3	J_4	SB	WB	SS	WS	J_1^*	J_2^{*MMC}	J_2^*	J_3^*	J_4^*	S^*	W^*
42-year period																
1/65–12/06	1.78	[3, 5]	17.86	17.61	18.01	1.39	1.45	4.95	6.61	4.32	[2, 5]	4.31	4.29	4.56	−0.50	0.13
	(0.06)	(0.07)	(0.05)	(0.06)	(0.05)	(0.84)	(0.79)	(0.89)	(0.76)	(0.04)	(0.05)	(0.04)	(0.04)	(0.03)	(0.61)	(0.90)
5-year subperiods and a 7-year subperiod																
1/65–12/69	2.02	[3, 21]	20.68	18.27	20.21	2.56	2.97	23.33	41.31	9.43	[1, 35]	9.04	8.66	9.53	1.46	1.98
	(0.05)	(0.07)	(0.02)	(0.05)	(0.03)	(0.10)	(0.03)	(0.01)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.14)	(0.05)
1/70–12/74	2.10	[2, 5]	21.41	18.91	19.61	2.56	2.52	40.80	39.05	0.96	[1, 6]	0.98	0.94	0.87	−2.56	−2.54
	(0.04)	(0.04)	(0.02)	(0.04)	(0.03)	(0.10)	(0.11)	(0.00)	(0.00)	(0.33)	(0.38)	(0.32)	(0.33)	(0.35)	(0.01)	(0.01)
1/75–12/79	1.94	[2, 5]	20.03	17.69	26.20	3.29	3.55	24.93	44.94	4.82	[1, 10]	4.79	4.59	5.46	1.10	1.90
	(0.06)	(0.07)	(0.03)	(0.06)	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.03)	(0.05)	(0.03)	(0.03)	(0.02)	(0.27)	(0.06)
1/80–12/84	1.17	[3, 23]	12.84	11.34	11.46	1.46	1.02	4.13	2.25	1.54	[1, 35]	1.57	1.51	1.58	−0.37	−0.50
	(0.33)	(0.36)	(0.23)	(0.33)	(0.32)	(0.79)	(0.98)	(0.94)	(0.99)	(0.22)	(0.25)	(0.21)	(0.22)	(0.21)	(0.72)	(0.61)
1/85–12/89	1.72	[4, 35]	18.08	15.97	15.35	1.83	1.53	10.93	7.20	3.04	[1, 35]	3.07	2.94	2.19	−1.46	−0.77
	(0.10)	(0.13)	(0.05)	(0.10)	(0.12)	(0.50)	(0.74)	(0.36)	(0.71)	(0.09)	(0.12)	(0.08)	(0.09)	(0.14)	(0.14)	(0.44)
1/90–12/94	1.09	[2, 5]	12.08	10.67	10.85	1.46	2.05	6.67	18.86	0.09	[1, 5]	0.10	0.09	0.10	0.73	1.53
	(0.39)	(0.40)	(0.28)	(0.38)	(0.37)	(0.79)	(0.34)	(0.76)	(0.04)	(0.76)	(0.84)	(0.76)	(0.76)	(0.76)	(0.46)	(0.12)
1/95–12/99	1.14	[2, 5]	12.58	11.11	12.51	2.19	1.88	8.27	9.62	1.09	[1, 35]	1.12	1.07	0.89	0.73	0.61
	(0.35)	(0.41)	(0.25)	(0.35)	(0.25)	(0.25)	(0.46)	(0.60)	(0.47)	(0.30)	(0.34)	(0.29)	(0.30)	(0.35)	(0.46)	(0.54)
1/00–12/06	1.90	[2, 5]	19.44	17.82	17.79	2.47	2.56	41.81	37.11	9.18	[1, 10]	8.91	8.65	9.01	1.23	1.56
	(0.06)	(0.05)	(0.03)	(0.06)	(0.06)	(0.13)	(0.10)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.22)	(0.12)
10-year subperiods and a 12-year subperiod																
1/65–12/74	2.33	[3, 5]	23.22	21.87	23.07	0.77	0.96	2.33	1.68	1.33	[2, 12]	1.35	1.32	1.27	0.26	-0.05
	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)	(0.99)	(0.98)	(0.99)	(0.99)	(0.25)	(0.27)	(0.25)	(0.25)	(0.26)	(0.80)	(0.96)
1/75–12/84	2.18	[2, 4]	21.87	20.60	25.51	3.10	3.20	60.67	55.14	6.82	[1, 6]	6.75	6.60	8.01	2.58	2.33
	(0.02)	(0.02)	(0.02)	(0.02)	(0.00)	(0.02)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.02)
1/85–12/94	1.90	[2, 5]	19.25	18.13	15.72	1.81	1.52	9.13	5.73	0.59	[1, 6]	0.60	0.59	0.53	0.77	0.16
	(0.05)	(0.06)	(0.04)	(0.05)	(0.11)	(0.52)	(0.75)	(0.52)	(0.84)	(0.44)	(0.58)	(0.44)	(0.44)	(0.47)	(0.44)	(0.87)
1/95–12/06	1.39	[2, 5]	14.34	13.65	13.51	3.30	2.99	36.38	21.23	3.34	[2, 14]	3.35	3.29	3.03	2.12	1.65
	(0.19)	(0.21)	(0.16)	(0.19)	(0.20)	(0.01)	(0.03)	(0.00)	(0.02)	(0.07)	(0.08)	(0.07)	(0.07)	(0.08)	(0.03)	(0.10)

Note: Test results are based on 10 size-based equity portfolios. The CRSP value-weighted index including distributions is used to proxy the market portfolio and a one-month Treasury bill is used as the riskless asset. The numbers in parentheses are p-values. The MMC tests are based on Student-t errors with unknown degrees of freedom. The square brackets are confidence intervals with level 0.975 for the degrees-of-freedom parameter and the MMC p-values are obtained by maximizing over the confidence interval.

Table 4
Estimated intercept values for individual portfolios and the aggregated one.

Time	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_A
42-year period											
1/65–12/06	7.19	3.83	2.74	2.52	1.97	2.52	2.26	1.29	1.23	−0.42	25.19
5-year subperiods and a 7-year subperiod											
1/65–12/69	26.00	19.07	12.44	12.02	10.48	7.62	6.46	4.44	3.03	−1.87	99.71
1/70–12/74	−3.82	−8.65	−7.61	−7.56	−7.09	−4.73	−3.10	−2.67	0.06	0.84	−44.35
1/75−12/79	15.46	14.22	13.64	14.29	10.89	11.63	9.56	6.31	3.68	−2.23	97.49
1/80–12/84	6.64	4.50	6.10	5.13	3.38	3.46	3.16	1.42	0.99	−0.85	33.96
1/85–12/89	−7.62	−7.52	−5.95	−5.91	−6.21	−3.90	−1.29	−0.88	−0.73	0.77	−39.27
1/90–12/94	6.64	2.35	−0.71	−0.97	1.62	1.28	0.76	0.46	1.08	−0.18	9.09
1/95–12/99	4.50	−0.96	−2.81	−4.27	−3.45	−4.78	−6.08	−6.19	−5.66	1.66	−28.05
1/00–12/06	9.65	7.39	8.59	8.85	10.01	9.48	8.28	6.88	5.88	−1.59	73.45
10-year subperiods and a 12-year subperiod											
1/65–12/74	12.27	6.18	3.29	2.96	2.33	1.99	2.14	1.13	1.60	−0.63	33.30
1/75–12/84	12.18	10.29	10.53	10.33	7.62	7.87	6.64	4.07	2.52	−1.62	70.46
1/85–12/94	−0.20	−2.43	−3.28	−3.36	−3.85	−1.25	−0.30	−0.17	0.15	0.28	−14.43
1/95–12/06	6.75	5.17	6.01	6.19	7.00	6.63	5.79	4.81	4.11	−1.11	51.39

Note: For each time period, the line entries are the estimated intercept in a regression of the portfolio excess returns on the market excess returns. The results for the individual portfolios are labeled by P_i and those for the aggregated one by $P_A = \sum_{i=1}^{10} P_i$. The intercept values are multiplied by 1000.

but all the other tests seem to favor a rejection. The results for the 10-year subperiods, reported in the bottom portion of Table 3, are more in line with those for the entire sample period. In the first three of those subperiods, the parametric tests tend to reject the null, while the distribution-free tests do not reject the null hypothesis of efficiency in two out of those four subperiods. Here the single-portfolio tests agree on rejections in the second and fourth subperiods, and non-rejections in the first and third subperiods.

Besides the obvious sensitivity to the choice of sample period, a comparison of the results across test statistics reveals that parametric versus non-parametric inference can differ markedly. Indeed, in many cases where the distribution-free tests indicate a clear non-rejection with large p-values, the parametric tests indicate the diametrical opposite with quite small p-values. The full sample results are a case in point.

Table 4 shows the OLS estimates of the intercept terms for each equation comprising the system in (2) and the equation in (14). The columns labeled P_1 to P_{10} show the estimates of a_i corresponding to the ten portfolios and the last column, labeled P_A, shows the estimates of the aggregate portfolio intercept $A = \sum_{i=1}^{10} a_i$. All the reported intercept values are multiplied by 1000. It is interesting to note that the individual intercept terms for portfolios P_1 to P_9 tend to all have the same sign, and the estimated intercept with the opposing sign for portfolio P_{10}—the largest cap-based decile portfolio—is quite small in magnitude compared to $\sum_{i=1}^{9} \hat{a}_i$. These results are evidence that the non-rejections by the distribution-free single-portfolio S^* and W^* tests seen in Table 3 are not due to the a_i's canceling out.

Table 5 reports the covariance matrix for the market model residuals, where the entries are multiplied by 1000. This matrix is an estimate of the covariances across portfolio returns once the effects of the market portfolio have been removed. The low covariance values suggest that assuming conditional independence in the cross-section of returns may be fairly innocuous in this application. Indeed, recall that the sum-type SS and WS tests assume that the error terms $(\varepsilon_{1t},..., \varepsilon_{Nt})$ are independent, conditional on R_p. The fact that those test statistics are not significant when computed with the full sample constitutes an implicit non-rejection of the conditional independence assumption. As a further check, we computed the Breusch and Pagan (1980) Lagrange multiplier test for the diagonality of the error covariance matrix of a seemingly unrelated equations system. The computed p-value of that test is 0.99, which clearly shows the joint statistical insignificance of the off-diagonal entries in Table 5. These results

Table 5
Covariance matrix of market model residuals.

Portfolio	1	2	3	4	5	6	7	8	9	10
1	4.16									
2	2.83	2.28								
3	2.20	1.74	1.55							
4	1.89	1.53	1.31	1.24						
5	1.63	1.33	1.15	1.06	1.01					
6	1.31	1.07	0.94	0.87	0.82	0.75				
7	0.98	0.82	0.73	0.69	0.65	0.58	0.53			
8	0.65	0.55	0.51	0.49	0.46	0.42	0.36	0.32		
9	0.35	0.30	0.28	0.28	0.27	0.26	0.24	0.20	0.18	
10	−0.25	−0.21	−0.18	−0.17	−0.16	−0.14	−0.12	−0.09	−0.06	0.03

Note: Entries are the covariances (multiplied by 1000) of the market model residuals on 10 size-based equity portfolios from January 1965 to December 2006. The CRSP value-weighted index including distributions is used to proxy the market portfolio and a one-month Treasury bill is used as the riskless asset.

support not only the validity of the sum-type *SS* and *WS* tests but also that of the single-portfolio S^* and W^* tests. In light of the simulation results, the most natural interpretation of the empirical findings is that the CRSP value-weighted index appears consistent with the mean-variance efficiency hypothesis over the full sample period.

7. Conclusion

The mean-variance efficiency of a given portfolio is typically assessed with tests based on OLS or GMM. The reliability of such test procedures depends on several parametric assumptions about the model's error terms, such as the finiteness and stationarity of moments. Exact parametric tests rely on an even stronger assumption about the actual distribution of the error terms. For instance, the well-known GRS Wald test assumes that the errors are multivariate Gaussian.

In this paper, we have proposed new non-randomized tests of mean-variance efficiency whose exactness does not depend on any parametric assumptions. The finite-sample validity of the new tests rests on the assumption that the joint temporal error density is symmetric around zero. This leaves open the possibility of conditional heteroskedasticity and even outliers. In fact, no restrictions are placed on the degree of non-normality or the degree of heterogeneity and dependence of the conditional variances. Not even the existence of moments is required.

The proposed tests were evaluated against commonly used parametric tests in some simulation experiments. The numerical results reveal that the parametric procedures suffer serious size distortions in the presence of contemporaneous conditional heteroskedasticity, especially when the number of test assets is large and/or the time history is relatively short. The MC approach of BDK can avoid this problem, but only if the precise form of heterogeneity is known. Since it has long been recognized that asset returns depart from homogeneous conditions, the new tests of the mean-variance efficiency hypothesis developed here offer a valid and useful distribution-free testing alternative to potentially very misleading parametric procedures.

Acknowledgements

We would like to thank two anonymous referees for their comments. We also thank Judith Clarke, Lynda Khalaf, Enrique Sentana, and Jay Shanken for useful discussions.

Appendix A

Proof of Proposition 1. Under Assumption 1, the elements of R_p are all different from zero with probability one, so the quantities ε_{it}/r_{pt}, $t=1,\ldots,T$, are well defined. Given R_p, we also have under Assumption 1 that

$$\begin{aligned}(\varepsilon_{i1}/r_{p1}, \varepsilon_{i2}/r_{p2}, \ldots, \varepsilon_{iT}/r_{pT}) &\overset{d}{=} (-\varepsilon_{i1}/r_{p1}, \varepsilon_{i2}/r_{p2}, \ldots, \varepsilon_{iT}/r_{pT}) \overset{d}{=} \ldots \\ &\overset{d}{=} (-\varepsilon_{i1}/r_{p1}, -\varepsilon_{i2}/r_{p2}, \ldots, -\varepsilon_{iT}/r_{pT}),\end{aligned} \tag{20}$$

where all 2^T such terms appear in this string of equalities in distribution. In particular, we have

$$(\varepsilon_{i1}/r_{p1}, \varepsilon_{i2}/r_{p2}, \ldots, \varepsilon_{iT}/r_{pT}) \overset{d}{=} (-\varepsilon_{i1}/r_{p1}, -\varepsilon_{i2}/r_{p2}, \ldots, -\varepsilon_{iT}/r_{pT}).$$

Define $\delta[\eta_1,\eta_2,\ldots,\eta_T] = ((\eta_{m+1}-\eta_1),(\eta_{m+2}-\eta_2),\ldots,(\eta_T-\eta_m))$. It follows that

$$\delta[\varepsilon_{i1}/r_{p1}, \varepsilon_{i2}/r_{p2}, \ldots, \varepsilon_{iT}/r_{pT}] \overset{d}{=} \delta[-\varepsilon_{i1}/r_{p1}, -\varepsilon_{i2}/r_{p2}, \ldots, -\varepsilon_{iT}/r_{pT}]$$

or

$$\begin{aligned}&((\varepsilon_{i,m+1}/r_{p,m+1}-\varepsilon_{i1}/r_{p1}), \ldots, (\varepsilon_{iT}/r_{pT}-\varepsilon_{im}/r_{pm})) \overset{d}{=} \\ &\quad(-(\varepsilon_{i,m+1}/r_{p,m+1}-\varepsilon_{i1}/r_{p1}), \ldots, -(\varepsilon_{iT}/r_{pT}-\varepsilon_{im}/r_{pm})),\end{aligned}$$

since $X \overset{d}{=} Y$ implies $U[X] \overset{d}{=} U[Y]$ for any measurable function $U[\cdot]$ defined on the common support of X and Y (Randles and Wolfe 1979, Theorem 1.3.7). This last result is used here conditional on R_p so that

$$\begin{aligned}&((\varepsilon_{i,m+1}/r_{p,m+1}-\varepsilon_{i1}/r_{p1})x_{p1}, \ldots, (\varepsilon_{iT}/r_{pT}-\varepsilon_{im}/r_{pm})x_{pm}) \overset{d}{=} \\ &\quad(-(\varepsilon_{i,m+1}/r_{p,m+1}-\varepsilon_{i1}/r_{p1})x_{p1}, \ldots, -(\varepsilon_{iT}/r_{pT}-\varepsilon_{im}/r_{pm})x_{pm}),\end{aligned}$$

since the x_{pt}'s defined in Eq. (7) are just functions of $(r_{p1},..., r_{pT})'$. In turn,

$$E[s[(\varepsilon_{i,t+m}/r_{p,t+m}-\varepsilon_{it}/r_{pt})x_{pt}]] = E[s[-(\varepsilon_{i,t+m}/r_{p,t+m}-\varepsilon_{it}/r_{pt})x_{pt}]]$$

or

$$Pr[(\varepsilon_{i,t+m}/r_{p,t+m}-\varepsilon_{it}/r_{pt})x_{pt}>0] = Pr[(\varepsilon_{i,t+m}/r_{p,t+m}-\varepsilon_{it}/r_{pt})x_{pt}<0] = 1/2, \tag{21}$$

for $t = 1, 2, ..., m$, since $Pr[(\varepsilon_{i,t+m}/r_{p,t+m}-\varepsilon_{it}/r_{pt})x_{pt} = 0] = 0$ under Assumption 2. Note that Eq. (21) holds conditional on R_p as well as unconditionally. □

Proof of Proposition 2. By applying the function $\delta[\cdot]$ defined in the proof of Proposition 1 to the string of equalities in distribution in Eq. (20), we see that:

$$\begin{aligned}
&((\varepsilon_{i,m+1}/r_{p,m+1}-\varepsilon_{i1}/r_{p1}),(\varepsilon_{i,m+2}/r_{p,m+2}-\varepsilon_{i2}/r_{p2}),...,(\varepsilon_{iT}/r_{pT}-\varepsilon_{im}/r_{pm})) \overset{d}{=}\\
&(-(\varepsilon_{i,m+1}/r_{p,m+1}-\varepsilon_{i1}/r_{p1}),(\varepsilon_{i,m+2}/r_{p,m+2}-\varepsilon_{i2}/r_{p2}),...,(\varepsilon_{iT}/r_{pT}-\varepsilon_{im}/r_{pm})) \overset{d}{=}\\
&((\varepsilon_{i,m+1}/r_{p,m+1}-\varepsilon_{i1}/r_{p1}),-(\varepsilon_{i,m+2}/r_{p,m+2}-\varepsilon_{i2}/r_{p2}),...,(\varepsilon_{iT}/r_{pT}-\varepsilon_{im}/r_{pm})) \overset{d}{=}\\
&(-(\varepsilon_{i,m+1}/r_{p,m+1}-\varepsilon_{i1}/r_{p1}),-(\varepsilon_{i,m+2}/r_{p,m+2}-\varepsilon_{i2}/r_{p2}),...,(\varepsilon_{iT}/r_{pT}-\varepsilon_{im}/r_{pm})) \overset{d}{=}\\
&\qquad\vdots\\
&\overset{d}{=}(-(\varepsilon_{i,m+1}/r_{p,m+1}-\varepsilon_{i1}/r_{p1}),-(\varepsilon_{i,m+2}/r_{p,m+2}-\varepsilon_{i2}/r_{p2}),...,-(\varepsilon_{iT}/r_{pT}-\varepsilon_{im}/r_{pm})),
\end{aligned}$$

where all 2^m such terms appear in this string of equalities in distribution. Let $E = ((\varepsilon_{i,m+1}/r_{p,m+1}-\varepsilon_{i1}/r_{p1})x_{p1},..., (\varepsilon_{iT}/r_{pT}-\varepsilon_{im}/r_{pm})x_{pm})$. It follows that the 2^m different values that the vector $s[E] = (s[(\varepsilon_{i,m+1}/r_{p,m+1}-\varepsilon_{i1}/r_{p1})x_{p1}],..., s[(\varepsilon_{iT}/r_{pT}-\varepsilon_{im}/r_{pm})x_{pm}])$ may take in $\{0,1\}^m$ have the same probability $(1/2)^m$. Therefore, the elements of $s[E]$ are mutually independent. Note that this result holds conditional on R_p as well as unconditionally. Define d_j to be the position of the integer j in the realization of the vector $(Rank(|z_{i1}|),..., Rank(|z_{im}|))$, $j = 1,...,m$, so that

$$\sum_{t=1}^{m} s[z_{it}]\varphi[Rank(|z_{it}|)] = \sum_{j=1}^{m} s[z_{i,d_j}]\varphi[j].$$

Conditional on $|E| = (|(\varepsilon_{i,m+1}/r_{p,m+1}-\varepsilon_{i1}/r_{p1})x_{p1}|,..., |(\varepsilon_{iT}/r_{pT}-\varepsilon_{im}/r_{pm})x_{pm}|)$, the vector of scores is a fixed permutation of $(\varphi[1],..., \varphi[m])$. So under the null hypothesis and conditional on $|E|$, it follows that

$$\sum_{j=1}^{m} s[z_{i,d_j}]\varphi[j] \overset{d}{=} \sum_{j=1}^{m} B_j\varphi[j],$$

since $(s[z_{i,d_1}],..., s[z_{i,d_m}]) \overset{d}{=} (B_1, B_2, ..., B_m)$, where $B_1,..., B_m$ are mutually independent uniform Bernoulli variables on $\{0,1\}$. Moreover, given the symmetry established in Proposition 1, we have under the null that $s[z_t]$ is independent of R_t^+ and thus of $\varphi[R_t^+]$ (Randles and Wolfe 1979, Lemma 2.4.2). Therefore, under the null, it is the case also unconditionally that

$$SR_i = \sum_{t=1}^{m} s[z_{it}]\varphi[Rank(|z_{it}|)] \overset{d}{=} \sum_{j=1}^{m} B_j\varphi[j],$$

since the distribution of $\sum_{j=1}^{m} B_j\varphi[j]$ does depend on $|E|$. □

References

Beaulieu, M.-C., Dufour, J.-M., Khalaf, L., 2007. Multivariate tests of mean-variance efficiency with possibly non-Gaussian errors: an exact simulation-based approach. Journal of Business and Economic Statistics 25, 398–410.

Berk, J., 1997. Necessary conditions for the CAPM. Journal of Economic Theory 73, 245–257.

Blattberg, R., Gonedes, N., 1974. A comparison of the stable and student distributions as statistical models of stock prices. Journal of Business 47, 244–280.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–327.

Bossaerts, P., Hillion, P., 1995. Testing the mean-variance efficiency of well-diversified portfolios in very large cross-sections. Annales d'Économie et de Statistique 40, 93–124.

Breusch, T.S., Pagan, A.R., 1980. The Lagrange multiplier test and its applications to model specification in econometrics. Review of Economic Studies 47, 239–253.

Campbell, J.Y., Lo, A.W., MacKinlay, A.C., 1997. The Econometrics of Financial Markets. Princeton University Press, New Jersey.

Chamberlain, G., 1983. A characterization of the distributions that imply mean-variance utility functions. Journal of Economic Theory 29, 185–201.

Chow, K.V., Denning, K.C., 1993. A simple multiple variance ratio test. Journal of Econometrics 58, 385–401.
Duffee, G.R., 1995. Stock returns and volatility: a firm-level analysis. Journal of Financial Economics 37, 399–420.
Duffee, G.R, 2001. "Asymmetric cross-sectional dispersion in stock returns: evidence and implications. Federal Reserve Bank of San Francisco Working Paper No. 2000-18.
Dufour, J.-M., Khalaf, L., 2001. Monte Carlo test methods in econometrics. In: Baltagi, B. (Ed.), Companion to Theoretical Econometrics. InBasil Blackwell, Oxford.
Dufour, J.-M., Khalaf, L., 2002. Simulation based finite and large sample tests in multivariate regressions. Journal of Econometrics 111, 303–322.
Fama, E., 1965. The behavior of stock market prices. Journal of Business 38, 34–105.
Gibbons, M.R., Ross, S.A., Shanken, J., 1989. A test of the efficiency of a given portfolio. Econometrica 57, 1121–1152.
Hollander, M., Wolfe, D.A., 1973. Nonparametric Statistical Methods. Wiley, New York.
Hsu, D.A., 1982. A Bayesian robust detection of shift in the risk structure of stock market returns. Journal of the American Statistical Association 77, 29–39.
Ingersoll, J., 1987. Theory of Financial Decision Making. Rowman & Littlefield.
Jobson, J.D., Korkie, B., 1982. Potential performance and tests of portfolio efficiency. Journal of Financial Economics 10, 433–466.
Lintner, J., 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. Review of Economics and Statistics 47, 13–37.
Lo, A.W., MacKinlay, A.C., 1990. Data-snooping biases in tests of financial asset pricing models. Review of Financial Studies 3, 431–467.
Luger, R., 2003. Exact non-parametric tests for a random walk with unknown drift under conditional heteroscedasticity. Journal of Econometrics 115, 259–276.
MacKinlay, A.C., Richardson, M.P., 1991. Using generalized method of moments to test mean-variance efficiency. Journal of Finance 46, 511–527.
Nelson, D.B., 1991. Conditional heteroskedasticity in asset returns: a new approach. Econometrica 59, 347–370.
Newey, W., West, K., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation covariance matrix. Econometrica 55, 703–708.
Owen, J., Rabinovitch, R., 1983. On the class of elliptical distributions and their applications to the theory of portfolio choice. Journal of Finance 38, 745–752.
Randles, R.H., Wolfe, D.A., 1979. Introduction to the Theory of Nonparametric Statistics. Wiley, New York.
Roll, R., 1979. A reply to Mayers and Rice (1979). Journal of Financial Economics 7, 391–400.
Savin, N.E., 1984. "Multiple hypothesis testing.". In: Griliches, Z., Intrilligator, M.D. (Eds.), Handbook of Econometrics Vol. 2. North-Holland, Amsterdam, pp. 827–879.
Sentana, E., 2008. "The econometrics of mean-variance efficiency tests: a survey." CEMFI Working Paper No. 0807.
Shanken, J., 1996. "Statistical methods in tests of portfolio efficiency: a synthesis.". In: Maddala, G.S., Rao, C.R. (Eds.), Handbook of Statistics, Vol. 14: Statistical Methods in Finance. North-Holland, Amsterdam.
Sharpe, W.F., 1964. Capital asset prices: a theory of market equilibrium under conditions of risk. Journal of Finance 19, 425–442.
Sidak, Z., 1967. Rectangular confidence regions for the means of multivariate normal distributions. Journal of the American Statistical Association 62, 626–633.
Theil, H., 1971. Principles of Econometrics. Wiley, New York.